VEN 009/01/PCT

JOINT INVENTORS

"EXPRESS MAIL" mailing label No. EL 485594042US
Date of Deposit: November 16, 2001
I hereby certify that this paper (or fee) is being deposited with the United States Postal Service 'EXPRESS MAIL POST OFFICE TO ADDRESSEE" service under 37 CFR § 1.10 on the date indicated above and is addressed to: Assistant Commissioner for Patents, Washington, D.C. 20231

Laura Handley

APPLICATION FOR UNITED STATES LETTERS PATENT

SPECIFICATION

TO WHOM IT MAY CONCERN:

Be it known that I, Carl Alexander Kamb, a citizen of the United States, residing at 696 Donner Hill Circle, Salt Lake City, in the County of Salt Lake and State of Utah; I, Sanghee Yoo, a citizen of Korea, I, Miguel Garcia-Guzman, a citizen of Spain, and I, Michael Leslie Pierce, residing at 2511 East 2860 South, Salt Lake City, in the county of Salt Lake and State of UT, have invented a new and useful "B-catenin pathway assays, and compositions therefrom," of which the following is a specification.

B-CATENIN ASSAYS, AND COMPOSITIONS THEREFROM

This application claims priority from and is a continuation-in-part of 08/812,994, now issued as U.S. Patent No. 5,955,275, and U.S. Application No. 60/253,325 (VEN009/00/P12, filed November 27, 2000), the entire disclosures of which are specifically incorporated by reference herein in their entireties.

FIELD OF THE INVENTION

The present invention relates to certain nucleic acid sequences, amino acid sequences, other compositions and methods relating to the characterization and physiologic implications of β-catenin/TCF pathways.

BACKGROUND OF THE INVENTION

Colorectal cancer is the second leading cause of cancer-related deaths in the United States, being responsible for as many as 60,000 fatalities each year. Nearly five percent of the US population develops colorectal cancer, and this number is predicted to rise as the average life expectancy increases (Beart, R.W. (1991) *American Cancer Society Textbook of Clinical Oncology. Atlanta*, American Cancer Society, pg. 213-218). Colon cancer erupts from the lumenal surface of the colon and rectum. Normally, epithelial cells line the surface of the colon and invaginate into structures called crypts. Over the course of three to six days, stem cells located at the base of each crypt divide, and then differentiate as they migrate toward the apex where they die and are released into the lumen (see, for example, Lipkin, M. et al. (1963) "Cell proliferation kinetics in the gastrointestinal tract of man." *J Clin Invest* 42:767). The first manifestations of colorectal cancer are often observed clinically as a polyp; a mass of epithelial cells that protrude from the apex of the colonic crypts of the bowel wall (see, for example, Kent, T.H. et al. (1983) "Polyps of the colon and small bowel, polyp syndromes, and the polyp carcinoma sequence." in Norris HT (eds) *Pathology of the Colon, Small Intestine, and Anus*. New York, Churchill Livingstone, vol 2, pg 167). Polyps are, predominantly, divided into two classes. The nondysplastic form consists of a large mass of cells that have normal morphology. These aggregates line up in a single row along the basement membrane and exhibit a low frequency of becoming neoplastic. The second form of polyp is the adenomatous polyp. These formations are dysplastic in nature and exhibit an abnormal intracellular and intercellular organization. As tumor progression evolves,

adenomatous polyps exhibit a high frequency of metastasis to surrounding tissues with the most common sites of invasion being the mesenteric lymph nodes, the peritoneal surface, and the liver.

Genetic studies have shown that at least two forms of heritable colorectal cancers exist. The first group, which includes familial adenomatous polyposis coli (FAP), Peutz-Jegher syndrome, familial juvenile polyposis, Cronkhite-Canada syndrome and hyperplastic polyposis, is characterized by the appearance of multiple (hundreds to thousands) of benign, precursor, colorectal polyps. In addition to colorectal lesions, several of these afflictions are associated with manifestations in other tissues including soft tissue tumors, osteomas, dental abnormalities, congenital hypertrophy of the retinal pigment epithelium (CHRPE), and cancers of the thyroid, small intestine, stomach, and brain (see, for example, Giardiello, F.M. (1995) "Gastrointestinal Polyposis syndromes and hereditary nonpolyposis colorectal cancer" in Rustgi AK (eds.) *Gastrointestinal Cancers:Biology, Diagnosis, and Therapy.* Philadelphia, Lippincott-Raven, pg. 367-377; Hamilton, S.R. et al. (1995) "The molecular basis of Turcot's syndrome." *New England J. Medicine* 332:839)). In contrast, patients with hereditary nonpolyposis colorectal cancer (HNPCC) lack an increase in the number of precursor adenomas yet share an increased risk for other ailments including cancer of the uterus, ovary and brain. Breakthroughs in molecular biology have identified several of the genes/gene families involved in the initiation and progression of colon cancer (see, for example, Vogelstein, B. et al. (1988) "Genetic alterations during colorectal-tumor development." *New England J. Medicine* 319:525; Kinzler, K. and Vogelstein, B. (1998) "Colorectal Tumors" in The Genetic Basis of Human Cancer. McGraw-Hill). In addition to identifying genes involved in DNA mismatch-repair (hMSH1, hMSH2, HPMS1, hPMS2), cell growth (e.g.the oncogenes K-ras, H-ras or N-ras), and cell cycle regulation (e.g tumor suppressors, p53), a growing body of evidence has suggested that a connection exists between cancer, cell adhesion and the Wnt/Wng pathway. Specifically attention has been drawn to a collection of gene products that include, but are not limited to, APC (adenomatous polyposis coli), β-catenin, TCF/LEF, GCK3, and cadherin.

The importance of APC in colorectal cancer was initially hypothesized when it was observed that roughly fifty percent of colorectal tumors exhibited cytologically

recognizable alterations of chromosome 5q, the position of the APC gene (see, for example, Kinzler, K.W. and Vogelstein, B. (1996) "Lessons from hereditary colorectal cancer." *Cell* 87:159-170). Since then, molecular analysis has shown that either APC or the gene encoding β-catenin are mutated at a high frequency in patients with FAP, sporadic colorectal tumors (see, for example, Ashton-Rickardt, P.G. (1989) "High frequency of APC loss in sporadic colorectal carcinoma due to breaks clustered in 5q21-22." *Oncogene* 4:1169; Sparks et al. (1998) " Mutational Analysis of the APC/β-catenin /TCF pathway in colorectal cancer." *Cancer Res.* 58:1130-1134) and other malignancies and diseases including, but not limited to, melanoma, hepatocellular carcinoma, ovarian cancer, endometrial cancer, medulloblastoma pilomatricomas, prostate cancer and Alzheimers (see, for example, Morin, P.J. (1999) "β-catenin signaling and cancer." *Bioessays* 21:1021-1030, Barker, N. et al. (2000) "The Yin-Yang of TCF/β-catenin Signaling." *Adv in Cancer Res.* 77:1-2; De Ferrari, G.V. and Inestrosa N.C. (2000) "Wnt signaling function in Alzheimer's disease." *Brain Res Brain Res Rev.* 33(1):1-12). Studies of the Wnt/Wng pathway have provided evidence for the role of APC in regulation of β-catenin. In the absence of a Wnt/Wng signal, a quaternary cytoplasmic complex comprising APC, β-catenin, glycogen synthase kinase-3β (GSK-3β) and conductin/axin is formed and mediates the phosphorylation and, consequently, targeted destruction of β-catenin via the ubiquitin proteasome pathway. Key to the formation of this complex is the association between APC and β-catenin; an interaction that is mediated by a 20 amino acid repeat found in the APC molecule (Figure 1, also, see, Su, L.K. et al. (1993) "Association of the APC tumor suppressor protein with catenins." *Science.*262(5140):1734-7). Natural mutations in this 20 a.a. sequence have been observed in many colon cancer lines and are associated with elevated levels of free β-catenin. In addition, it has been observed that reintroduction of wild type APC into these cells results in a dramatic down regulation of the free, cytoplasmic form of β-catenin thus suggesting that APC plays a critical role in the modulation of free, cytoplasmic β-catenin (Munemitsu, S. et al (1995) "Regulation of intracellular β-catenin levels by the adenomatous polyposis coli (APC) turmor-suppressor protein." *PNAS* 92:3046-3050).

APC's ability to down-regulate cytoplasmic β-catenin levels is dependent on its interaction with two negative regulators of the Wnt/Wng pathway: GSK-3β and conductin. Biochemical studies have shown that GSK-3β phosphorylates both β-catenin and APC, and that these events lead to an increase in APC's affinity to β-catenin and the eventual destruction of the molecule (Munemitsu, S. et al (1995)). Further evidence for the role of GSK-3β in β-catenin stability comes from the observation that loss of GSK-3β activity leads to increased levels of cytoplasmic β-catenin (He, X. et al. (1995) "Glycogen synthase kinase-3 and dorsoventral patterning in Xenopus embryos." *Science* 374(6523):617-22). Conductin's interaction with β-catenin was initially identified by two-hybrid interaction assays. Subsequently, this molecule was shown to have additional binding sites for APC and GSK-3β and is now believed to act as a docking molecule or platform for the APC, β-catenin, GSK-3β complex.

While the absence of a Wnt/Wng signal leads to the destruction of β-catenin, activation of the pathway counteracts the negative regulation of β-catenin. In response to the Wnt/Wng signal, the GSK 3β- binding protein (GBP) binds to GSK-3β and prevents phosphorylation of the relevant targets, APC and β-catenin. As a result of these events, cytoplasmic levels of free β-catenin are elevated, thus promoting the interaction between β-catenin and a family of transcription factors called TCF's (T-cell factors).

The connection between cell adhesion and cancer has been speculated upon for several years. Tumor cells making the transition from benign lesions to malignant, invasive cancers, need to overcome cell-cell adhesion in order to invade surrounding tissues. Several studies have observed that the expression of cell adhesion molecules, including E-cadherin, are lost or altered during the development of many human cancers (Birchmeier W. and Behrens J. (1994) "Cadherin expression in carcinomas: role in the formation of cell junctions and the prevention of invasiveness." *Biochim Biophys Acta.* 1198(1):11-26) and that co-expression of a dominant negative form of E-cadherin can result in the development of adenomas in mice (see, for example, Michelle, L. et al. (1995) "Inflamatory Bowel Disease and Adenomas in Mice Expressing a Dominant Negative N-Cadherin." *Science* 270:1203-1207). In many instances, forced reintroduction of wildtype adhesion molecules into invasive tumors results in the reversion to the

benign, non-invasive phenotype (Frixen ,U.H. et al. (1991) "E-cadherin-mediated cell-cell adhesion prevents invasiveness of human carcinoma cells." *J Cell Biol.* 113(1):173-85) again suggesting a strong correlation between cancer and cell adhesion.

The bridge between APC/β-catenin and cell adhesion comes from both histological observations and molecular data. Studies in wild-type colorectal epithelial cells have shown that APC expression gradually increases as cells migrate to the top of the colonic crypt, and peaks just prior to when the cells undergo apoptosis and slough off into the lumen of the colon (see, Miyashiro, I. et al (1995) "Subcellular localization of the APC protein: Immunoelectron microscopic study of the association of the APC protein with catenin." *Oncogene* 11:89). Furthermore, loss of just a single copy of the APC gene results in a decrease of the enterocyte crypt-to-villus migration (Mahmoud, N.N. et al. (1997) "Apc gene mutation is associated with a dominant-negative effect upon intestinal cell migration." *Cancer Res* 57(22):5045-50). Since loss of cells from the top of crypts is an important homeostatic process, it has been suggested that disruptions in cellular adhesion via mutations in APC and APC-interactive proteins may lead to alterations in normal cell growth regulation and/or apoptosis (see, for example, Wijnhoven B.P. (2000) "E-cadherin-catenin cell-cell adhesion complex and human cancer." *Br J Surg.* 87(8):992-1005; Kim, K. et al, (2000) "Overexpression of beta-Catenin Induces Apoptosis Independent of Its Transactivation Function with LEF-1 or the Involvement of Major G1Cell Cycle Regulators." *Mol Biol Cell.* 11(10):3509-3523; Weihl, C.C. (1999) "The role of beta-catenin stability in mutant PS1-associated apoptosis." *Neuroreport* 10(12):2527-32). Evidence supporting such a model come from the discovery that β-catenin associates with the cytoplasmic tail of cadherins (see Kemler, R. (1993) "Cytoplasmic protein interactions and regulation of cell adhesion." *Trends in Genetics* 9:317). Studies have shown that over-expression of C-cadherin alters the Wnt/Wng signaling cascade and that these alterations are mediated through the region of cadherin that interacts with β-catenin (see, for instance, Fagotto, F. et al. (1996) "Binding to Cadherins Antagonizes the Signaling Activity of β-catenin during Axis Formation in Xenopus." *J of Cell Biology*, 132:1105-1114). Given that binding of β-catenin to APC and cadherins is mutually exclusive (both bind to a set of repeated

elements called "armadillo repeats") it is possible that APC/β-catenin fulfill some facet of their tumor suppressor function by modulating cell adhesion in this fashion.

Despite detailed knowledge of these and other genes involved in colorectal cancer, the art to date has not provided an efficient method for exploring the biological intricacies of colon cancer and identifying new putative therapeutic drugs for the prevention and treatment of this disease. Prophylactic colectomies are still routinely performed on FAP patients as a preferred method to reduce the risk of cancer and patients with metastatis disease usually receive radiation and/or current, broad-acting chemotherapeutic agents. Although such treatments can induce temporary remissions, they are often not curative, as evidenced by the fact that approximately 40% of the colon cancer patients die from the disease within 5 years. The present invention provides an opportunity to identify new drugs and drug targets that can be utilized to battle the increasing incidence of colon cancer that is predicted for the upcoming decade.

BRIEF SUMMARY OF THE INVENTION

The present invention relates to activity of β-catenin-related pathways, as well as to compositions therefrom. More specifically, the present invention generally relates to methods for assessing β-catenin pathway-related activity, and from such methods, obtaining perturbagens with β-catenin-related activity. Such perturbagens then are used to obtain β-catenin-related targets, which in turn can be used to identify potential therapeutics. The invention also provides genetic material for the development of gene therapy agents, vectors and host cells.

The present invention provides polypeptides of cadherin perturbagens V, VI and XI, biologically active fragments, analogs and modifications thereof, and polypeptides consisting essentially of such perturbagen sequences. In other aspects, the invention provides polypeptides having at least 99%, at least 95%, at least 90%, at least 85% or at least 80% sequence identitity or homology with such perturbagens, and in other aspects provides N- and C-terminal fragments of such perturbagens. The invention further provides a composition of such polypeptides in a pharmaceutically acceptable carrier, and for treating a β-catenin-related condition with a therapeutically effective amount of a polypeptide of the invention.

The present invention also provides polypeptides having β-catenin pathway activity that are fused to heterologous sequences, in some aspects a scaffold or more particularly, a fluorescent protein scaffold, and provides polypeptides having β-catenin pathway activity that are chemically modified, or more particularly, radiolabelled, acetylated, glycosylated, or fluorescently tagged. Antibodies to the polypeptides of the invention also are provided.

The present invention further provides polynucleotides encoding cadherin perturbagens V, VI and XI, biologically active fragments, analogs and modifications thereof, and polypeptides consisting essentially of such perturbagen sequences. In other aspects, the invention provides polynucleotides encoding polypeptides having at least 99%, at least 95%, at least 90%, at least 85% or at least 80% sequence identity or homology with such perturbagens, and in other aspects provides polynucleotides encoding N- and C-terminal fragments of such perturbagens. In some aspects, the polynucleotides are chemically synthesized.

The present invention further provides host cells, vectors, and gene therapy vectors comprising the polynucleotides of the invention. The host cells of the invention further provide for methods for producing β-catenin-related polypeptides by culturing such host cells and recovering such polypeptides.

The present invention also provides methods for identifying a cellular target that interacts with the polypeptides of the invention. In some aspects, the method is performed *in vitro* and comprises detecting reporter expression, and in particular aspects, utilizes a yeast two-hybrid assay format. The present invention further provides for the use of such target in screening for putative β-catenin-related therapeutics, and in some aspects screens for disruption of polypeptide-target pairs. In particular aspects, a combinatorial chemical library is so screened.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure Legends

Figure 1. The Wnt/Wng pathway and proposed interactions with cadherin. The interaction of the Wnt/Wng molecule with its receptor leads to activation of GBP (GSK-3β binding protein) and inhibition of β-catenin degradation by the ubiquitin pathway. As a result, cytoplasmic β-catenin levels increase, leading to formation of the B-catenin-Tcf

complex and heightened transcription of genes carrying a Tcf binding element. Increased levels of free β-catenin also lend themselves to B-catenin-is rapidly degraded via the ubiquitin pathway.

Figure 2. A. Mapping the functional region of a perturbagen. Four perturbagens are derived from different breakpoints within the same gene. By mapping the smallest sequence that is common to all four perturbagens (dotted line) it is possible to identify biologically critical regions (black box). B. Critical regions of a gene can be determined by deletion analysis. For instance, a series of N-terminal deletions (dotted line) can be tested for biological activity. In this example, full activity requires a molecule that is longer than deletion 2 but smaller than deletion 1.

Figure 3. Isolation of a β-catenin/Tcf reporter line. A population of cells containing the TBE-GFP reporter construct and the dominant allele of β-catenin (β-cat S45Y) undergoes multiple rounds of FACS to enrich for bright (GFP^+) cells. Subsequently, cells are plated at low density to isolate individual clones. Samples of each clone are then transduced with the dominant negative allele of Tcf4 (Tcf4 Δ30). Cell lines that are responsive to the dominant effects of Tcf4 Δ30 are identified by FACS and the parent clone is recovered by returning to the original plate (i.e. stripped wells).

Figure 4. A perturbagen library is introduced into the S4535 clone and screened by FACS to isolate dim clones. Genomic DNA prepared from the dim cells is then used to PCR amplify the perturbagen encoding inserts in each cell. Each insert is then subcloned into the original vector (pVT352.1) and reinfected into a fresh population of S4535 cells for further enrichment.

Figure 5. Basic two-hybrid methodology. When bait and prey molecules interact, the Gal4-AD and Gal40-BD binding domains of the Gal4 transcriptional activator are reconstituted. As a result, this functional unit can associate with the Gal1 UAS and induce transcription of the reporter gene (Leu2).

Figure 6. Four-Hybrid System. Host cell RNA targets are identified through a four-hybrid modification of the original two-hybrid scheme. Expanded region (lower left) pictures interaction between "bait" and "target" RNA molecules.

Figure 7. LANCE™. In the homogeneous assay, a Cy5 labeled perturbagen binds to an Eu-Target molecule in solution. A. When the two molecules are in close proximity, the

emissions of the lanthanide chelate can excite Cy5 and give rise to a robust signal. B. In the presence of a small molecule inhibitor, the Cy5-perturbagen-Target-Eu interaction is prevented. Subsequent excitation of Eu results in little or no signal.

Figure 8. DELFIA™. In the heterogeneous assay, the target is immobilized to a solid support using an Eu labeled monoclonal antibody. Following incubation with the Cy5 labeled perturbagen, the well is washed to remove unbound Cy5. Due to the close proximity of the Eu and Cy5 moieties in the bound complex, excitation of the lanthanide chelate leads to excitation (and emission) of Cy5. In the presence of a small molecule inhibitor (black circles), the Eu-target and Cy5-perturbagen moieties never come in close proximity. In subsequent washes, the free, unbound, Cy5-peptide conjugate is removed and the Eu induced Cy5 signal is insignificant.

Figure 9. Construction of the full length β-catenin clone.

Figure 10. Construction of the full length Tcf clone.

Figure 11. Histogram comparing the fluorescent distribution of i) HEK293 cells, ii) HEK293 cells with the TBE2 x 4 –GFP reporter, iii) HEK293 cells with the TBE2 x 4 – GFP reporter and β-cat S37F, iv) HEK293 cells with the TBE2 x 4 –GFP and β-cat S45. The y-axis represents cell number. The x-axis represents fluorescence intensity.

Figure 12. Responsiveness of Clone S4535. A. Histogram of HEK 293 cells. B. Histogram of Clone S4535. C. Clone S4535 with Tcf 4 Δ30.

Figure 13 -15. Cadherin Perturbagens. The DNA and peptide sequences of perturbagens are listed. In some cases a second DNA sequence indicates the reverse strand (labeled R) of the perturbagen insert. Penetrance for each clone is also displayed. Isoforms of cadherin are identified.

Figure 16. Histogram of S4535 cells containing CadV.

Figure 17. Bar Graph showing effects of TcfDN and Cad5CD on dead cell number in HT29 cells, HEK293 cells, HMEK cells, and HUVEC's.

Figure 18. Bar Graph showing effects of TcfDN and Cad5CD on total cell number in HT29 cells, HEK293 cells, HMEC cells, and HUVEC's.

Figure 19. Histogram of cells used in gene profiling studies.

Figure 20. List of target genes and/or Est sequences that are altered in both Tcf4DN and Cad5CD transformed lines. The "ratio" refers to numbers derived from calculations made

by Incyte Genomics. "Z" values were determined using calculations described in Kamb et al.

Figure 21. List of target genes and/or Est sequences that are altered in Cad5CD transformed lines.

Figures 22-24. Diagrams of vectors used in these studies.

DEFINITIONS

The terms "perturbagen" or "phenotypic probe" refers to an agent that is proteinaceous or ribonucleic in nature and acts in a transdominant mode to interfere with specific biochemical processes in cells, i.e., through its interaction with specific cellular target(s) or other such component(s), capable of disrupting or activating a particular signaling pathway and/or cellular event. Perturbagens may be encoded by a naturally derived library of compounds such as a cDNA or genomic DNA (gDNA) expression library, or an artificial library comprising synthetic oligonucleotide sequences of a desired length or range of lengths, e.g. a random peptide library. Alternatively, the perturbagen itself can be synthesized using chemical methods. The term "proteinaceous perturbagen" encompasses peptides, oligo- or polypeptides, proteins, protein fragments, or protein variants. Some proteinaceous perturbagens can be as short as three amino acids in length. Alternatively, these agents can be greater than 3 amino acids but less than ten amino acids. Other agents can be greater than ten amino acids but shorter than 30 amino acids in length. Still other agents can be greater than 30 amino acids but less than 100 amino acids in length. Still other agents can be greater than 100 amino acids in length. Naturally occurring proteinaceous perturbagens (i.e. those derived from cDNA or genomic DNA) exhibit a range in size from as little as three to several hundred amino acids. In contrast, synthetic perturbagens (such as those present in a synthetic peptide library) may range in size from three amino acids to fifty amino acids in length and more preferably, from three to 20 amino acids in length, and yet more preferably, about 15 amino acids in length. Similarly, the length of RNA perturbagens can vary. Some RNA perturbagens are as short as 6-10 nucleotides in length. Other RNA perturbagens are between 10 and 50 nucleotides in length. Still other RNA perturbagens are between 50 and 200 nucleotides in length. Other RNA perturbagens are greater than 200 nucleotides in length.

The term "mimetic" refers to a small molecule that (i) exerts the same or similar physiological or phenotypic effect in a bioassay system or in an animal model as does a given perturbagen, or (ii) is capable of displacing a perturbagen from a target in a displacement assay.

The term "small molecule" refers to a chemical compound, for instance a peptide or oligonucleotide that may optionally be derivatized, natural product or any other low molecular weight (less than about 1 kDalton) organic, bioinorganic or inorganic compound, of either natural or synthetic origin. Such small molecules may be a therapeutically deliverable substance or may be further derivatized to facilitate delivery.

The term "target" refers to any cellular component that is directly acted upon by the perturbagen that leads to and/or induces the phenotypic change, detectible for example in a bioassay system.

The terms "library" or "genetic library" refer to a collection of nucleic acid fragments that may individually range in size from about a few base pairs to about a million base pairs, with typical expression libraries of about nine base pairs to about ten thousand base pairs. These fragments are generated using a variety of techniques familiar to the art.

The term "sublibrary" refers to a portion of a genetic library that has been isolated by application of a specific screening or selection procedure.

The term "insert" in the context of a library refers to an individual DNA fragment that constitutes a single member of the library.

The terms "reporter gene" and "reporter" refer to nucleic acid sequences or encoded polypeptides for which screens or selections can be devised. Reporters may be proteins capable of emitting light, or genes that encode intracellular or cell surface proteins detectable by antibodies. Preferably, the reporter activity may be evaluated in a quantitative manner. Alternatively, reporter genes can confer antibiotic resistance or selectable growth advantages.

The term "gene" refers to a DNA substantially encoding an endogenous cellular component, and includes both the coding and antisense strands, the 5' and 3' regions that are not transcribed but serve as transcriptional control domains, and transcribed but

untranslated domains such as introns (including splice junctions), polyadenylation signals, ribosomal recognition domains, and the like.

The terms "polynucleotide" or "nucleic acid molecule" are used interchangeably to refer to polymeric forms of nucleotides of any length. The polynucleotides may contain deoxyribonucleotides, ribonucleotides and/or their analogs. Nucleotides may have any three-dimensional structure, and may perform any function, known or unknown. The term "polynucleotide" includes single-, double-stranded and triple helical molecules. "Oligonucleotide" refers to polynucleotides of between 5 and about 100 nucleotides of single- or double-stranded DNA. Oligonucleotides are also known as oligomers or oligos and may be isolated from genes, or chemically synthesized by methods known in the art. The following are non-limiting embodiments of polynucleotides: a gene or gene fragment, exons, introns, mRNA, tRNA, rRNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes and primers. A nucleic acid molecule may also comprise modified nucleic acid molecules, such as methylated nucleic acid molecules and nucleic acid molecule analogs. Analogs of purines and pyrimidines are known in the art, and include, but are not limited to, aziridinycytosine, 4-acetylcytosine, 5-fluorouracil, 5-bromouracil, 5-carboxymethylaminomethyl-2-thiouracil, 5-carboxymethyl-aminomethyluracil, inosine, N6-isopentenyladenine, 1-methyladenine, 1-methylpseudouracil, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, pseudouracil, 5-pentylnyluracil and 2,6-diaminopurine. The use of uracil as a substitute for thymine in a deoxyribonucleic acid is also considered an analogous form of pyrimidine.

The term "fragment" refers to any portion of a proteinaceous perturbagen that is at least 3 amino acids in length, or any RNA molecule that is at least 5 nucleotides in length. The descriptors "biologically relevant" or "biologically active" refer to that portion of a protein or protein fragment, RNA or RNA fragment, or DNA fragment that encodes either of the two previous entities, that is responsible for an observable phenotype, some portion of an observable phenotype, or for activation of a correlative reporter construct.

The term "variant" refers to biologically active forms of the perturbagen sequence (or the polynucleotide sequence that encodes the perturbagen) that differ from the sequence of the initial perturbagen.

The terms "homology" or "homologous" refers to the percentage of residues in a candidate sequence that are identical with the residues in the reference sequence after aligning the two sequences and introducing gaps, if necessary, to achieve the maximum percent of overlap (see, for example, Altschul, S.F. et al. (1990) "Basic local alignment search tool." *J Mol Biol* 215(3):403-10; Altschul, S.F. et al. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Res* 25(17):3389-402). It is understood that homologous sequences can accommodate insertions, deletions and substitutions in the nucleotide sequence. Thus, linear sequences of nucleotides can be essentially identical even if some of the nucleotide residues do not precisely correspond or align. The reference sequence may be a subset of a larger sequence, such as a portion of a gene or flanking sequence, or a repetitive portion of a chromosome.

The term "scaffold" refers to a proteinaceous or RNA sequence to which the perturbagen is covalently linked to provide e.g., conformational stability and/or protection from degradation.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Agents isolated from the methods described herein have broad potential and application. Among others, the invention permits the definition of disease pathways, the identification of diagnostically or therapeutically useful targets, and the identification of therapeutic agents. For example, β-catenin/TCF pathway-related genes that are mutated, up-regulated, or down-regulated under disease conditions may be involved in causing or exacerbating the disease condition. Treatments directed at modulating the activity of such genes or treatments that involve alternate pathways may ameliorate the disease condition. Also, the agents and assays described herein thus have utility as models for diseases related to the β-catenin/TCF pathway. The assays may be utilized as part of a screening strategy designed to identify agents such as compounds that are capable of ameliorating disease symptoms.

As more fully disclosed herein, the described methodology yields first a set of RNA-based or proteinaceous agents, and second, a set of endogenous cellular targets. Each RNA-based or proteinaceous agent (or a mimetic, agonist or antagonist thereof identified through, e.g., routine small molecule screens) may be useful as a direct therapeutic agent in the treatment of cancer and/or various diseases. With each such new agent, a corresponding target molecule can be readily identified using standard interaction methodologies such as the two-hybrid technique. Such targets are useful in the development of novel drugs for new chemotherapeutic strategies and may provide useful diagnostic tools for profiling the genetic background (genotype) of the particular disease under study.

A. Overview of the Invention

The invention describes the isolation of new and previously unidentified agents that alter the sensitivity of a cell to the activation of the βcatenin/TCF pathways. The perturbagens described herein were isolated using a phenotypic assay. See priority document U.S. Patent No. 5,955,275, "Methods for identifying nucleic acid sequences encoding agents that affect cellular phenotypes," the disclosure of which is incorporated by reference herein in its entirety. Briefly, the assay identifies agents that alter a cell's responsiveness to an activated form of β-catenin (e.g. β-catenin S45T, or β-catenin S37F). To accomplish this, a library of polynucleotide sequences is generated using a variety of techniques familiar to the art. After ligating this material into a standard expression vector, the library is transferred into a population of cells of a given type (e.g. a cell line) and screened for sequences that induce a particular biological phenotype. The assay advantageously identifies one or more relevant sequences from the library in the selected host cell population. Cells expressing a biologically relevant perturbagen induce a particular phenotype (or correlative activation of a reporter gene), and are then separated from the rest of the population using, e.g., high-throughput Fluorescent Activated Cell Sorting (FACS) screening procedures. Such high-throughput machines are both highly sensitive and efficient (obtaining screening speeds of approximately 10,000 to approximately 65,000 cells or more per minute) thus facilitating identification of biologically relevant sequences that exist at low frequencies within a cell population.

Here, an assay has been designed to identify molecules that alter a cell's ability to respond to a mutated form of β-catenin (β-catenin Serine45Tyrosine, also referred to herein as β-catenin S45Y or β-cat S45Y). To accomplish this, a random primed library constructed from cDNA was transfected into a subline of the HEK 293 (human embryonic kidney) cell line that i) expressed the dominant, constitutively active form of β-catenin (β-catenin S45Y) and ii) contained a transcriptionally regulated reporter construct consisting of four tandem TCF-4 binding elements (TBE's) functionally linked to the coding region of EGFP. Subsequently, roughly 20 million of these cells, representing a 2x fold coverage of the library, were then subjected to FACS analysis to identify perturbagens that disrupted the β-catenin S45Y driven activation of the reporter construct.

Perturbagen identification may elucidate the function of known genes, or alternatively may work in a "black-box" approach to identify new genes, gene products, or cellular targets. Thus in some instances, perturbagens may be encoded by a previously identified gene (or gene fragment thereof). Such a gene may be one whose contribution to the disease pathway has previously been identified. Alternatively, the contribution of a gene to the pathway may have been previously unrecognized. In yet other cases, the perturbagen may be found to have no homology with any previously identified polynucleotide or proteinaceous agent. Such perturbagens may be derived from previously unidentified genes, or alternatively may be random sequences that have the proper conformation and/or chemical characteristics needed to alter or modulate one or more components of a pathway(s) that influences the phenotype under investigation. In the methodology described herein, no prior knowledge of the perturbagen or of its corresponding gene, gene product or cellular target is necessary. Moreover, because it is possible for multiple perturbagens to assume similar secondary or tertiary conformations and/or have shared or related chemistries, two or more variants of the same perturbagen may be identified and isolated from a single library without any additional screening steps. Thus one need not spend laborious hours designing, redesigning, or manipulating any candidate molecules, and thus does not bias the experiment with preconceived conceptions of what will or will not induce the phenotype of interest.

B. Phenotypic Probes

The invention encompasses both the phenotypic probes (perturbagens) described herewith and the polynucleotide sequences encoding them. As one of ordinary skill appreciates, such agents may be described by their RNA sequence, amino acid sequence, or correlative DNA sequence. Alternatively, the agents can be sufficiently described in terms of their identity as isolates of a library that exhibit a particular biological activity.

Perturbagens may be encoded by a variety of genetic libraries, including those developed from cDNA, gDNA, and random, synthetic oligonucleotides synthesized using current available methods in chemistry (see, for example, Caponigro et al. (1998) "Transdominant genetic analysis of a growth control pathway." *PNAS* 95:7508-7513; Caruthers, M.H. et al. (1980) Nucleic Acids Symposium, Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symposium, Ser. 7:225-232; Cwirla, S.E. et al. (1990) "Peptides on phage: a vast library of peptides for identifying ligands." *Proc Natl Acad Sci* 87(16):6378-82). Alternatively, the perturbagen itself can be synthesized using chemical methods. For example, peptide and RNA synthesis can be performed using various techniques (Roberge, J.Y. et al. (1995) "A strategy for a convergent synthesis of N-linked glycopeptides on a solid support." *Science* 269:202-204; Zhang, X. et al. (1997) "RNA synthesis using a universal base-stable allyl linker." *NAR* 25(20): 3980-3983) and diverse combinatorial peptide libraries can be constructed using, a variety of strategies such as the multipin strategy, the tea bag method", or the split-couple-mix method (see, for instance, Geysen, H.M. et al (1984) "Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acids." *PNAS* 81:3998-4002; Houghten, R.A. (1985) "General methods for the rapid solid phase synthesis of large numbers of peptides: specificity of antigen-antibody interaction at the level of individual amino acids." *PNAS* 82:5131-5135; Lam, K.S. et al. (1991) "A new type of synthetic library for identifying ligand binding activity." Nature 354:82-84; Al-Obeidi, F. et al. (1998) "Peptide and Peptidomimetic Libraries." Molecular Biotechnology: 9:205-223). Automated synthesis may be achieved using commercially available equipment such as the ABI 431A peptide synthesizer (Perkin-Elmer).

In some cases, the polynucleotide sequence encoding a perturbagen represents a fragment of an existing gene. Using currently available software, it is possible to identify

the full length cDNA by aligning the perturbagen encoding sequence with pre-existing sequences maintained in, for instance, publicly available genomic and/or EST data bases. In situations where the gene has not been identified, the perturbagen can be readily used to reverse engineer and identify the gene from which the phenotypic probe is derived.

In the case where a perturbagen is encoded by only a portion of a particular gene, the nucleic acid sequence of such a perturbagen may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences. One such method, restriction site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector (Sarkar, G. (1993) "Restriction-site PCR: a direct method of unknown sequence retrieval adjacent to a known locus by using universal primers." *PCR Methods Applic.* 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (see Triglia, T. et al. (1988) "A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences." *NAR.* 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) "Capture PCR: efficient amplification of DNA fragments adjacent to a known sequence in human and YAC DNA." *PCR Methods Applic.* 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double stranded sequence into a region of known sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al (1991) "Targeted gene walking polymerase chain reaction." *NAR.* 19:3055-3060). In addition, one may use nested primers and PROMOTERFINDER libraries (Clontech, Palo Alto, CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR based methods, primers may be designed, using commercially available software such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

In one particular embodiment, the invention encompasses proteinaceous perturbagens, biologically active fragments, (N-terminal, C-terminal, or internal) or variants thereof. Proteinaceous perturbagens can exert their effects by multiple means. For example, a peptide may act by binding and disrupting the interactions between two or more proteinaceous entities within the cell. Alternatively, a peptide perturbagen can bind to, and disrupt translation of a particular mRNA molecule. As still another alternative, peptide perturbagens may bind to genomic DNA and disrupt gene expression by altering the ability of one or more transcription factor(s) (e.g. activators or repressors) from binding to a critical enhancer/promoter region of the regulatory region of the gene.

Penetrance is another property of perturbagens. Penetrance is defined as the number of cells exhibiting a particular phenotype divided by the total number of cells in the experiment (when a perturbagen is present in the cells), minus the total number of cells exhibiting a particular phenotype divided by the total number of cells in the experiment when the perturbagen is not present in the cells. The penetrance of any given pertubagen can vary depending upon a variety of parameters including 1) the cell type it is being expressed in, 2) the vector being used to express the perturbagen, 3) the biological stability (half-life) of the perturbagen or mRNA encoding the perturbagen 4) the concentration of the perturbagen in the cell, as well as other parameters. Thus although penetrance is a factor that impacts how immediately a given perturbagen can be seen to exert an effect, in some instances, a desirable, biologically active perturbagen may present a relatively low rate of penetrance. As one of ordinary skill will appreciate, perturbagens of low penetrance may be obtained and manipulated via standard cycling and/or amplification procedures. Thus, some preferred perturbagens may exhibit as low as 1-2% penetrance. Other preferred perturbagens may exhibit between 2% and 5% penetrance, between 5 and 10% penetrance, 10% and 20% penetrance, between 20% and 50% penetrance, or even in some instances, between 50% and 100% penetrance.

In some instances, the action, penetrance, or biological activity of a perturbagen may be affected in some part by the scaffold to which it is associated. In some cases (for instance, in situations where the agent is shorter than 30 amino acids) the scaffold may drive the perturbagen to adopt a conformation that enhances its biological action. In still other instances, one or more neighboring residues from, e.g., the C-terminus of a

scaffold, may act in concert with the perturbagen to enhance the functionality of the molecule. In cases such as these, the complete biologically active sequence may include one or more C-terminal residues derived from the scaffold molecule. Multiple techniques may be used to determine the contribution of the scaffold to the phenotypic effect of any given perturbagen. Initially, perturbagen sequences can be shifted to alternative scaffolds and retested for biological activity. If these procedures result in a significant loss of the perturbagen's activity, a fusion between the perturbagen and, for instance, the 30-most residues from the C-terminus of the scaffold may be linked to a second scaffold molecule and retested for biological activity. Should operations such as these lead to the recovery of lost activity, experiments in which smaller and small portions of the primary scaffold are associated with the perturbagen can be tested.

In other embodiments, the phenotypic probe is an RNA molecule which is itself active (i.e. is not acting through the correlative encoded protein or peptide that results from translation of the RNA). There are multiple mechanisms by which RNA molecules may act to inhibit or activate a biological pathway. In some instances, the RNA perturbagen acts in an antisense mode to disrupt ribonucleic acid transcription or translation of a cellular mRNA target via hybridization to a target ribonucleic acid (Weiss, B. et al. (1999) "Antisense RNA gene therapy for studying and modulating biological processes." *Cell Mol Life Sc.i* 55(3):334-58). In this context the term "antisense" refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a particular target DNA (see, for example, Chadwick, D.R. et al. (2000) "Antisense RNA sequences targeting the 5' leader packaging signal region of human immunodeficiency virus type-1 inhibits viral replication at post-transcriptional stages of the life cycle." *Gene Therapy* 7(16):1362-8). In other instances, RNA perturbagens may act as a RNA-PRO agents, disrupting β-catenin-TCF pathway by interacting with one or more proteinaceous components (e.g. APC) of the cell (see Sengupta, D.J. (1999) "Identification of RNAs that bind to a specific protein using the yeast three-hybrid system." *RNA* 5:596-601). In still other instances, RNA perturbagens may act as a triplex-forming oligonucleotide (TFO) agent to interact with promoter sequences, exons, introns, or other portions of genomic DNA to disrupt or activate transcription of components of the β-catenin-TCF pathway (see Postel,

E.H. et al. (1989) "Evidence that a triplex-forming oligonucleotide binds to the c-myc promoter in HeLa cells, thereby reducing c-myc RNA levels." *PNAS* 88: 8227-8231; Svinarchuk, F. et al. (1997) "Recruitment of transcription factors to the target site by triplex-forming oligonucleotides." *NAR* 25:3459-3464).

There does not appear to be a necessary correlation between size of a particular RNA (or proteinaceous) perturbagen and penetrance. Instead, the penetrance of perturbagens are dependent upon the perturbagen stability or half-life, the perturbagen's ability to achieve access to the target molecule, and other factors.

Perturbagens may also exhibit cross-reactivity. A variety of host target proteins can contain similarities in both the primary and secondary structure. As a result, one or more of the agents described herein may exhibit affinity for one or more target variants/isoforms present in nature. Similarly, agents identified in the following screens may exhibit affinity for two or more functionally unrelated proteins that contain regions or domains that share homology or related functional groups. Thus, for instance, a perturbagen that recognizes a zinc-binding domain of one protein, may also show affinity for the homologous (and functionally equivalent) region of a second protein (see, e.g., Mavromatis K. O. et al. (1997) "The carboxyl-terminal zinc-binding domain of the human papillomavirus E7 protein can be functionally replaced by the homologous sequences of the E6 protein." *Viral Research* 52(1):109-18). In cases where such interactions lead to relevant biological phenotypes, the underlying mechanism(s) may differ considerably from those brought about by the original perturbagen-target interactions. Furthermore, in cases where an agent exhibits cross reactivity with secondary targets, said agents may be useful in a broader set of therapeutic and diagnostic applications than originally intended.

Host range is another characteristic of perturbagens. The term "host range" refers to the breadth of potential host cells that exhibit perturbagen-induced phenotypes. In some instances, such as the case where the perturbagen is represented by an apoptosis-inducing fragment of BID, the host range is broad, due to the near ubiquitous participation of BID or BID-like agents in the apoptotic pathway of many cells. In contrast, some perturbagens have a very limited host range due to, for instance, the restricted expression of the perturbagen target.

C. Sequence Variants

In another embodiment, the invention includes sequence variants of both the phenotypic probes and the polynucleotide sequences that encode them. Thus, in the case of proteinaceous perturbagens, variants contain at least one amino acid substitution, deletion, or insertion from the original isolated form of the perturbagen that provides biological properties that are substantially similar to those of the initial perturbagen. Similarly, variants of RNA-based phenotypic probes contain at least one nucleotide substitution, deletion, or insertion when compared to the original isolated sequence.

In addition to being described by their respective sequence, variants may also be identified by the relative amounts of homology they have in common with the original perturbagen sequence. Alternatively, a variant of a proteinaceous perturbagen may be described in terms of the nature of an amino acid substitution. "Conservative" substitutions are those in which the substituting residue is structurally or functionally similar to the substituted residue. In non-conservative substitutions, the substituting and substituted residue will be from structurally or functionally different classes. For the purposes herein, these classes are as follows: 1. Electropositive: R, K,H; 2. Electronegative: D,E; 3. Aliphatic: V,L,I,M; 4. Aromatic: F,Y,W; 5. Small: A,S,T,G,P,C; 6. Charged: R,K,D,E,H; 7. Polar: S,T,Q,N,Y,H,W; and Small Hydrophilic: C,S,T. Interclass substitutions generally are characterized as nonconservative, while intraclass substitutions are considered to be conservative.

In some instances, variant polypeptides sequences can have 65-75% homology with the original agent. In other embodiments, variants have between 75% and 85% homology with the original agent. In still other embodiments, variants will have between 85% and 95% homology with the original perturbagen agent. In yet other embodiments, variants have between 95% and greater than 99% polypeptide sequence identity with the original perturbagen agent. In some cases, the homology between two perturbagens (variants) is confined to a small region of the molecule (e.g. a motif). Such conserved sequences are often indicative of regions that contain biologically important functions and suggest the perturbagens share a common cellular target. In these situations, while only limited and conservative amino acid changes are desirable within the region of the

motif, greater levels of variation can exist in adjacent and more distal portions of the polypeptide.

Like their proteinaceous counterparts, variants of RNA perturbagens may also be described in terms of percent homology. In some instances, the variant ribonucleotide sequences can have 65-75% homology with the original agent. In other embodiments, the variants have between 75% and 85% homology with the original agent or between 85% and 95% homology with the original perturbagen sequence, or even between 95% and greater than 99% sequence identity with the original perturbagen agent. Again, greater variation can, in some embodiments, exist outside an identified region/motif without altering biological activity.

Lastly, in reference to the DNA sequences encoding proteinaceous perturbagens, one who is skilled in the art will appreciate that the degree of variance will depend upon and/or reflect the degeneracy of the genetic code. As one in the art appreciates, a given protein sequence is equivalently encoded by a large number of polynucleotide sequences. Therefore, the invention encompasses each variation of polynucleotide sequence that encodes the given perturbagen, such variations being made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of each perturbagen. For each proteinaceous perturbagen described by amino acid sequence herein, all such corresponding DNA variations are to be considered as being specifically disclosed.

Variants of phenotypic probes may arise by a variety of means. Some variants may be artifactual and result from, for instance, errors that occur in the process of PCR amplification or cloning of the perturbagen encoding sequence. Alternatively, variants may be constructed intentionally. For instance, it may be advantageous to produce nucleotide sequences encoding perturbagens possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide or RNA occurs in a particular prokaryotic or eukaryotic cell in accordance with the frequency with which particular codons are utilized by the host (Berg, O.G. (1997) "Growth rate-optimized tRNA abundance and codon usage." *J Mol Biol* 270(4):544-50). Additional reasons for substantially altering the nucleotide sequence encoding proteinaceous perturbagens (without altering the encoded amino acid sequences) include,

but are not limited to, producing RNA transcripts that have increased half-life. This may be accomplished by altering a sequence's structural stability (see, for example, Gross, G. et al. (1990) "RNA primary sequence or secondary structure in the translational initiation region controls expression of two variant interferon-beta genes in Escherichia coli." *J Biol Chem.* 265(29):17627-36; Ralston, C.Y. et al. (2000) "Stability and cooperativity of individual tertiary contacts in RNA revealed through chemical denaturation." *Nat Struct Biol.* 7(5):371-4), or through addition of untranslated sequences that increase RNA stability/half-life through RNA-protein interactions (see, for example, Wang ,W. et al. (2000) "HuR regulates cyclin A and cyclin B1 mRNA stability during cell proliferation." *EMBO J.* 19(10):2340-50; Shetty, S. and Idell, S. (2000) "Posttranscriptional regulation of plasminogen activator inhibitor-1 in human lung carcinoma cells in vitro." *Am J Physiol Lung Cell Mol Physiol* 278(1):L148-56). Also included the category of intentional variants are those whose sequence has been altered in order to add or deleted sites involved in post-translational modification. Included in this list are variants in which phosphorylation sites, acetylation sites, methylation sites, and/or glycosylation sites have been added or deleted (see, for example, Wicker-Planquart, C. (1999) "Site-directed removal of N-glycosylation sites in human gastric lipase." *Eur J Biochem.* 262(3):644-51; Dou, Y. (1999) "Phos-phorylation of linker histone H1 regulates gene expression in vivo by mimicking H1 removal." *Mol Cell.* 4(4):641-7).

Variants may also arise as a result of simple and relatively routine techniques involving random mutagenesis or "DNA shuffling"; procedures that are often used to rapidly evolve perturbagen encoding sequences and allow identification of variants that have increased biological stability or activity (see, for instance, Ner, S.S. et al. (1988) "A simple and efficient procedure for generating random point mutations and for codon replacements using mixed oligonucleotides." *DNA* 7:127-134; Stemmer ,W. (1994) "Rapid evolution of a protein in vitro by DNA shuffling." *Nature* 370:389-391). For instance, in mutagenic PCR, the fragment encoding the perturbagen is PCR amplified under conditions that increase the error rate of *Taq* polymerase. In general, this is accomplished by i) increasing the $MgCl_2$ concentrations to stabilize non-complementary pairings, ii) addition of $MnCl_2$ to diminish template specificity of the polymerase and iii) increasing the concentration of dCTP and dTTP to promote misincorporation of basepairs

in the reaction. As a result of this process, the error rate of *Taq* polymerase may be increased from 1.0 x 10^{-4} errors per nucleotide per pass of the polymerase, to approximately 7 x 10^{-3} errors per nucleotide per pass. Amplifying a perturbagen-encoding sequence under these conditions allows the development of a library of dissimilar sequences which can subsequently be screened for variants that exhibit improved biological activity.

In addition to variants that are created by artificial or accidental means, natural variants may also exist. For instance, in the course of screening any given genomic or cDNA library, it is possible that a perturbagen, derived from a sequence that exists in multiple copies within the genome (e.g. duplications, repetitive sequences), may be isolated numerous times. Such sequences often contain polymorphisms that result in alterations in the encoded RNA and polypeptide sequence (see, for example, Satoh, H. et al. (1999) "Molecular cloning and characterization of two sets of alpha-theta genes in the rat alpha-like globin gene cluster." *Gene* 230(1):91-9) and thus, may represent natural variants of the perturbagen agent. Alternatively, if multiple libraries are utilized to screen for perturbagens and two or more of those libraries are derived from unrelated individuals, dissimilar tissues, or contrary periods in the development of a tissue (e.g. adult vs. fetal tissue) it is possible that variants may be isolated as a result of allelic variation (see, for example, Posnett, D.N. (1990) "Allelic variations of human TCR V gene products." *Immunol Today*. 11(10):368-73). Variants of phenotypic probes may arise by these and other means.

Variants of any given perturbagen may in some instances exhibit additional biological properties. For instance, perturbagens that previously recognized only a single target may demonstrate broadened specificity, e.g., may bind multiple isoforms or serotypes of a target in response to the alteration of a single amino acid in the perturbagen variant. Similarly, a perturbagen having a specific phenotype in one cell may exhibit additional phenotypes or may exhibit a broader effective host range after making small alterations in the perturbagen variant sequence.

D. Biologically Active Fragments

Some embodiments of the invention encompass "biologically active fragments" of a given proteinaceous or RNA-based perturbagen. Biologically active fragments may be

compromised of N-terminal, C-terminal, or internal fragments of peptide perturbagens, or 5', 3'or internal fragments of RNA perturbagens. In some instances, the fragment encodes or represents portions of a natural gene. In other instances the fragment is derived from a larger polynucleotide or polypeptide that has no known natural counterpart. In still other instances, biologically active regions of a perturbagen can be artificially synthesized (by chemical or recombinant methods) so that multiple, tandem copies of the phenotypic probe are covalently linked together and expressed. All such biologically active perturbagen fragments are, in turn, encoded by a variety of correlative DNA sequences.

The biologically active portion of a molecule can be identified by several means. In some instances, biological relevant regions can be deduced by simple physical mapping of families of overlapping sequences isolated from a phenotypic assay (Hingorani, K. et al. (2000) "Mapping the functional domains of nucleolar protein B23." *J Biol Chem* May 26). For instance, in the course of any given screen, multiple perturbagens, derived from alternative breakpoints of the same gene, may be isolated from one or more genetic libraries. (Figure 2). The smallest region that is common to all of the perturbagens can demarcate the area of biological importance.

Alternatively, critical regions of a perturbagen can frequently be distinguished by comparing the polynucleotide and/or amino acid sequence of two or more perturbagens that share a common target (see, for example, Grundy, W.N. (1998) "Homology detection via family pair-wise search." *J Comput Biol.* 5(3):479-9; Gorodkin, J. et al. (1997) "Finding common sequence and structure motifs in a set of RNA sequences." *Ismb* 5:120-3). In this instance, conserved sequences (or motifs) that are identified by this form of analysis often provide important clues necessary to determine biologically important regions of a given molecule. Alternatively, methods that identify biologically relevant regions by altering or deleting regions of the perturbagen molecule can also be used. For instance, the gene encoding a particular perturbagen can be subjected to deletion analysis whereby portions of the gene are removed in a systematic fashion, thus allowing the remaining entity to be retested for its ability to evoke a biological response (see, for example, Huhn, J. et al. (2000) "Molecular analysis of CD26-mediated signal transduction in cells." *Immunol Lett* 72(2):127-132; Davezac, N. et al. (2000)

"Regulation of CDC25B phosphatases subcellular localization." *Oncogene* 19(18):2179-85).

Alternatively, biologically critical regions of a molecule can be identified by inducing mutations in the sequence encoding the polypeptide (see, for example, Ito, Y. et al. (1999) "Analysis of functional regions of YPM, a superantigen derived from gram-negative bacteria." *Eur J Biochem*; 263(2):326-37; Kim, S.W. et al. (2000) "Identification of functionally important amino acid residues within the C2-domain of human factor V using alanine-scanning mutagenesis." *Biochemistry* 39(8):1951-8.). Subsequent testing of the variants of said molecule for biological activity enables the investigator to identify regions of the perturbagen that are both critical and sensitive to manipulation. Furthermore, molecular probes such as monoclonal antibodies and epitope-specific peptides can be useful in the identification of biologically important regions of a perturbagen (see, for example, Midgley, C.A. et al. (2000) "An N-terminal p14ARF peptide blocks Mdm2-dependent ubiquitination in vitro and can activate p53 in vivo." *Oncogene* 19(19):2312-23; Lu, D. et al. (2000) "Identification of the residues in the extracellular region of KDR important for interaction with vascular endothelial growth factor and neutralizing anti-KDR antibodies." *J Biol Chem* 275(19):14321-30). In this procedure, probes that bind and thus mask specific regions of a perturbagen can be tested for their ability to block the biological activity of the molecule. These techniques (as well as others) can be used to map the boundaries of any given biologically active residues.

E. Heterologous Sequences

In another embodiment, the invention encompasses all heterologous forms of the phenotypic probes and the polynucleotide sequences encoding them described herewith. In this context, "heterologous sequence(s)" include versions of the perturbagens that are i) scaffolded by other entities, ii) tagged with marker sequences that can be recognized by antibodies or specific peptides, iii) altered to transform post-translational patterns of modification or iv) altered chemically so as to cyclize the molecule for alternative pharmacodynamic/pharmacokinetic properties.

1. Scaffolds

Peptide perturbagens can be fused to protein scaffolds at N-terminal, C-terminal, or internal sites. Similarly, RNA derived perturbagens can be fused to RNA sequences at 5', 3'or internal sites. The fusion of a perturbagen to a second entity can increase the relative effectiveness of the perturbagen by increasing the stability of either the messenger RNA (mRNA) or protein of said agent. In some instances, scaffolds may be a relatively inert protein, (i.e. having no enzymatic activity or fluorescent properties) such as hemagglutinin. Such proteins can be stably expressed in a wide variety of cell types without disrupting the normal physiological functions of the cell. In other instances, scaffolds may serve a dual function, e.g., increasing perturbagen stability while at the same time, serving as an indicator or gauge of the level of perturbagen expression. In this case, the scaffold may be an autofluorescent molecule such as a green fluorescent protein (Clontech) or embody an enzymatic activity capable of altering a substrate in such a way that it can be detected by eye or instrumentation (e.g. β galactosidase). For example, in the invention described herein, various molecular techniques that are common to the field are used to link the perturbagen library to, e.g., the C-terminus of a nonfluorescent variant of GFP. "dEGFP" (also referred to as "dead-GFP") is one such nonfluorescent variant brought about by conversion of Tyr → Phe at codon 66 of EGFP (Clontech). By linking the perturbagen library to this molecule, each library member is fused to a separate dEGFP molecule. Such chimeric fusions can easily be detected by Western Blot analysis using antibodies directed against GFP and are useful in determination of intracellular expression levels of perturbagens. In addition, by modifying the perturbagen sequences or the scaffold to which they are attached with various localization signals, the perturbagen may be directed to a particular compartment within the host cell. For example, proteinaceous perturbagens can be directed to the nucleus of certain cell types by attachment of a nuclear localization sequence (NLS); a heterogeneous sequence made up of short stretches of basic amino acid residues recognized by importins alpha and/or beta.

2. Antibody-Tagged Perturbagens

Perturbagens can be constructed to contain a heterologous moiety (a "tag") that is recognized by a commercially available antibody. Such heterologous forms may

facilitate studies of subjects including, but not limited to, i) perturbagen subcellular localization, ii) intracellular concentration assessment and iii) target binding interactions. In addition, the tagging of a perturbagen may also facilitate purification of fusion proteins using commercially available matrices (see, for example, James, E.A. et al. "Production and characterization of biologically active human GM-CSF secreted by genetically modified plant cells." *Protein Expr Purif.* 19(1):131-8; Kilic, F. and Rudnick, G. (2000) "Oligomerization of serotonin transporter and its functional consequences." *Proc Natl Acad Sci U S A.* 97(7):3106-11). Such moieties include, but are not limited to glutathione-S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, c-myc, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, c-myc and HA enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. Such fusion proteins may also be engineered to contain a proteolytic cleavage site located between the perturbagen sequence and the heterologous protein/tag sequence, so that the perturbagen may be cleaved away from the heterologous moiety following purification. A variety of commercially produced kits may be used to facilitate expression and purification of fusion proteins.

3. Chemically Modified Perturbagens

In addition to the chimeric variants described above, chemical modification encompass a variety of modifications including, but not limited to, perturbagens that have been radiolabeled with ^{32}P or ^{35}S, acetylated, glycosylated, or labeled with fluorescent molecules such as FITC or rhodamine. These modifications may be directly imposed on the perturbagen itself (see, for example, Shuvaev, V.V. et al. (1999) "Glycation of apolipoprotein E impairs its binding to heparin: identification of the major glycation site." *Biochim Biophys Acta* 1454(3):296-308; Dobransky, T. et al. (2000) "Expression, purification and characterization of recombinant human choline acetyltransferase: phosphorylation of the enzyme regulates catalytic activity." *Biochem J.* 349(Pt 1):141-151). Alternatively, changes may be made to the polynucleotide sequence encoding the perturbagen so as to alter the pattern of phosphorylation, acetylation, glycosylation, or

that lead to cyclization of peptides in order to alter membrane permeability and/or pharmacodynamic- pharmacokinetic properties (see, for example, Borchardt, R.T. (1999) "Optimizing oral adsorption of peptides using prodrug strategies." J Control Release 62(1-2):231-8.).

F. Hybridization

The invention also encompasses polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences encoding phenotypic probes and said variants of such entities described previously, under various conditions of stringency. Such reagents may be useful in i) therapeutics, ii) diagnostic assays, iii) immunocytology, iv) target identification, and v) purification. For example, if the sequence encoding a particular perturbagen is introduced into a subject for gene therapeutic purposes, it may be necessary to monitor the success of integration and the levels of expression of said agent by Southern and Northern Blot analysis respectively (Pu, P. et al. (2000) "Inhibitory effect of antisense epidermal growth factor receptor RNA on the proliferation of rat C6 glioma cells in vitro and in vivo." *J Neurosurg.* 92(1):132-9). In other instances, hybridization may be used as a tool to define or describe a perturbagen variant or fragment, and a hybridizing sequence thus may have direct relevance as a mimetic or other such therapeutic agent.

The term "hybridization" refers to any process by which a strand of nucleic acid binds with a complementary or near-complementary strand through base pairing. There are several parameters that play a role in determining whether two polynucleotide molecules will hybridize including salt concentrations, temperature, and the presence or absence of organic solvents. For instance stringent salt concentrations will ordinarily be less than about 750mM NaCl and 75mM trisodium citrate, preferably less than about 500mM NaCl and 50mM trisodiium citrate, and most preferably less than about 250mM NaCl and 25mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent (e.g. formamide) while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30°C, more preferably of at least about 37°C, and most preferably of at least about 42°C. Varying additional parameters, such as hybridization

time, the concentration of detergent and the inclusion or exclusion of carrier DNA are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30°C in 750 mM NaCl, 75mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37°C in 500 mM NaCl, 50mM trisodium citrate, 1% SDS, 35% formamide and 100ug/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42°C in 250 mM NaCl, 25mM trisodium citrate, 1% SDS, 50% formamide and 200ug/ml denatured ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The washing steps that follow hybridization can also vary greatly in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentrations for the wash steps will preferably be less than about 30mM NaCl and 3mM trisodium citrate, and most preferably less than about 15mM NaCl and 1.5mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include temperatures of at least about 25°C, more preferably of at least about 42°C, and most preferably of at least about 68°C. In a preferred embodiment, wash steps will occur at 25°C in 30mM NaCl, 3mM trisodium citrate and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42°C in 15mM NaCl, 1.5mM trisodium citrate and 0.1% SDS. In a most preferred embodiment, wash steps will occur at 68°C in 15mM NaCl, 1.5mM trisodium citrate and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art.

G. Expression Vectors

The DNA sequence encoding each perturbagen or target (or variant or fragment thereof) may be inserted into an expression vector which contains the necessary elements for transcriptional/translational control in a selected host cell. Thus the DNA sequence may be expressed for, e.g., testing in a bioassay such as those described herein, or in a binding assay such as those described herein, or for production and recovery of the proteinaceous agent. Methods which are well known to those skilled in the art are used

to construct expression vectors containing sequences encoding the perturbagens and the appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination (see Sambrook, J. et al. (1989) "Molecular Cloning, A Laboratory Manual", Cold Spring Harbor Press, Plainview NY).

Exemplary expression vectors may include one or more of the following: (i) regulatory sequences, such as enhancers, constitutive and inducible promoters, and/or (ii) 5′ and 3′ untranslated regions, and/or (iii) mRNA stabilizing sequences or scaffolds, for optimal expression of the perturbagen in a given host. For instance, intracellular perturbagen levels can be modulated using alternative promoter sequences such as CMV, RSV, and SV40 promoters, to drive transcription (see, for example, Zarrin, A.A. et al. (1999) "Comparison of CMV, RSV, SV40 viral and Vlambda1 cellular promoters in B and T lymphoid and non-lymphoid cell lines." *Biochim Biophys Acta.* 1446(1-2):135-9). Alternatively, inducible promoter systems, (e.g. ponesterone-induced promoter (PIND, Invitrogen, see Dunlop, J. et al. (1999) "Steroid hormone-inducible expression of the GLT-1 subtype of high-affinity l-glutamate transporter in human embryonic kidney cells." *Biochem Biophys Res Commun.* 265(1):101-5), tissue specific enhancers (see Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162), or scaffolding molecules (see, for example, see Abedi, M. et al. (1998), "Green fluorescent protein as a scaffold for intracellular presentation of peptides." *Nucleic Acid Research* 26(2):623-630) can be used to modulate intracellular perturbagen levels.

A variety of paired expression vector/host systems may be utilized to contain and express sequences encoding the perturbagens. As one of ordinary skill will appreciate, the selection of a given system is dictated by the purpose of expression: e.g., bioassay, binding assay, or production of proteinaceous product for subsequent isolation and purification. Such systems include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors, insect cell systems infected with viral expression vectors (e.g. baculovirus), plant cell systems transformed with viral expression vectors (e.g. tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g. Ti or pBR322 plasmids; or mammalian cell systems (*e.g.* COS,

CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (*e.g.*, metallothionine promoter) or from mammalian viruses (*e.g.*, the adenovirus late promoter; the vaccinia virus 7.5 K promoter). The host cell employed does not limit the invention.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding the perturbagens. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding perturbagens can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla Ca). Ligation of sequences encoding perturbagens into the vector's cloning site disrupts the lacZ gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (see e.g., Van Heeke, G. and Schuster, S.M. (1989) "Expression of human asparagine synthetase in Escherichia coli." *J. Biol. Chem.* 264:5503-5509). When large quantities of perturbagens are needed, e.g. for the production of antibodies, vectors which direct high level expression of perturbagens may be used. Exemplary vectors feature the strong, inducible T5 or T7 bacteriophage promoter; the *E. coli* expression vector pUR278 (Ruther *et al.*, *EMBO J.*, 2:1791-94 (1983)), in which the gene protein coding sequence may be ligated individually into the vector in frame with the *lac Z* coding region so that a fusion protein is produced; pIN vectors (Inouye & Inouye, *Nucleic Acids Res.*, 13:3101-09 (1985); Van Heeke *et al.*, *J. Biol. Chem.*, 264:5503-9 (1989)); and the like. pGEX vectors may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned anaphylatoxin C3a receptor gene protein can be released from the GST moiety.

Yeast expression systems may also be used for production of perturbagens. A number of vectors containing constitutive or inducible promoters such as alpha factor,

alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerivisiae or related strains. In addition, such vectors can be designed to direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences in the host genome for stable propagation. (see, e.g. Bitter, G.A. et al. (1987) "Expression and secretion vectors for yeast." *Methods Enzymology*. 153:516-544; and Scorer, C.A. et al. (1994) "Rapid selection using G418 of high copy number transformants of Pichia pastoris for high-level foreign gene expression." *Biotechnology* 12:181-184).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the gene coding sequence of interest may be ligated to an adenovirus transcription/translation control complex, *e.g.*, the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (*e.g.*, region E1 or E3) will result in a recombinant virus that is viable and capable of expressing gene protein in infected hosts. (*e.g.*, *see* Logan *et al.*, *Proc. Natl. Acad. Sci. USA*, 81:3655-59 (1984)). Specific initiation signals may be used to achieve more efficient translation of sequences encoding the perturbagen. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding the perturbagen and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence is inserted, exogenous translational control signals including an in-frame ATG initiation codon are provided by the vector. Furthermore, the initiation codon must be in phase with the reading frame of the desired coding sequence to ensure translation of the entire insert. Such exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (see Bitter, *et al., Methods in Enzymol.,* 153:516-44 (1987)). Alternatively, many of these elements are not required in vectors that are specific for RNA-based perturbagens. Instead, sequences that stabilize the RNA transcript or direct the RNA sequence to a particular compartment will be

included (see, for instance, Wood Chuck post transcriptional regulatory element, WPRE, Zufferey, R. et al. (1999) "Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors." *J Virol* 73(4):2886-92).

Plant systems may also be used for expression of perturbagens. Transcription of sequences encoding perturbagens may be driven by viral promoters, e.g. the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1991) "Deletion analysis of the 5' untranslated leader sequence of tobacco mosaic virus RNA." *J Virology* 65:1619-22). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (see, for example, Coruzzi, G. et al. (1984) "Tissue-specific and light-regulated expression of a pea nuclear gene encoding the small subunit of ribulose-1,5-bisphosphate." *EMBO J.* 3:1671-80; Broglie, R. et al. (1984) "Light-regulated expression of a pea ribulose-1,5-bisphosphate carboxylase small subunit gene in transformed plant cells." *Science* 24:838-843).

In an insect system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in Spodoptera frugiperda cells. The gene coding sequence may be cloned individually into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of gene coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (*i.e.*, virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted gene is expressed (*see, e.g.*, Smith, *et al.*, *J. Virol.* 46: 584-93 (1983); U.S. Patent No. 4,745,051).

In addition, a host cell strain may be chosen that modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (*e.g.*, glycosylation) and processing (*e.g.*, cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins. Appropriate cell lines or host systems can be chosen to ensure

the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells that possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. Such mammalian host cells include but are not limited to CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, etc.

The selected construct can be introduced into the selected host cell by direct DNA transformation or pathogen-mediated transfection. The terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Preferred technologies for introducing perturbagens into mammalian cells include, but are not limited to, retroviral infection as well as transformation by EBV or similar episomally-maintained viral vectors (Makrides, S.C. (1999) "Components of vectors for gene transfer and expression in mammalian cells." *Protein Expr Purif* 17(2):183-202). Other suitable methods for transforming or transfecting host cells can be found in Maniatis, T. et al ("Molecular Cloning: A Laboratory Manual." Cold Spring Harbor Laboratory Press) and other standard laboratory manuals.

For long term production of recombinant proteins in mammalian systems, stable expression of perturbagens in cell lines is preferred. For example, sequences encoding perturbagens can be transformed or introduced into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Alternatively, cells can be transfected using, for instance, retroviral, adenoviral, or adeno-associated viral agents as delivery systems for the perturbagen. For example, retroviral vectors (e.g. LRCX, Clontech) may be used to introduce and express perturbagens in a variety of mammalian cell cultures. Such vectors may rely on the virus' own 5' LTR as a means of driving perturbagen expression or may utilize alternative promoters/enhancers (e.g. those of CMV, RSV and SV40, PIND) to regulate perturbagen expression levels.

In a preferred embodiment, timing and/or quantity of expression of the recombinant protein can be controlled using an inducible expression construct. Inducible constructs and systems for inducible expression of recombinant proteins will be well

known to those skilled in the art. Examples of such inducible promoters or other gene regulatory elements include, but are not limited to, tetracycline, metallothionine, ecdysone, and other steroid-responsive promoters, rapamycin responsive promoters, and the like (No, *et al.*, *Proc. Natl. Acad. Sci. USA*, 93:3346-51 (1996); Furth, *et al.*, *Proc. Natl. Acad. Sci. USA*, 91:9302-6 (1994)). Additional control elements that can be used include promoters requiring specific transcription factors such as viral, particularly HIV, promoters. In one in embodiment, a Tet inducible gene expression system is utilized. (Gossen *et al.*, *Proc. Natl. Acad. Sci. USA*, 89:5547-51 (1992); Gossen, *et al.*, *Science*, 268:1766-69 (1995)). Tet Expression Systems are based on two regulatory elements derived from the tetracycline-resistance operon of the *E. coli* Tn10 transposon—the tetracycline repressor protein (TetR) and the tetracycline operator sequence (*tetO*) to which TetR binds. Using such a system, expression of the recombinant protein is placed under the control of the *tetO* operator sequence and transfected or transformed into a host cell. In the presence of TetR, which is co-transfected into the host cell, expression of the recombinant protein is repressed due to binding of the TetR protein to the *tetO* regulatory element. High-level, regulated gene expression can then be induced in response to varying concentrations of tetracycline (Tc) or Tc derivatives such as doxycycline (Dox), which compete with *tetO* elements for binding to TetR. Constructs and materials for tet inducible gene expression are available commercially from CLONTECH Laboratories, Inc., Palo Alto, CA.

When used as a component in an assay system, the gene protein may be labeled, either directly or indirectly, to facilitate detection of a complex formed between the gene protein and a test substance. Any of a variety of suitable labeling systems may be used including but not limited to radioisotopes such as ^{125}I; enzyme labeling systems that generate a detectable calorimetric signal or light when exposed to substrate; and fluorescent labels. Where recombinant DNA technology is used to produce the gene protein for such assay systems, it may be advantageous to engineer fusion proteins that can facilitate labeling, immobilization and/or detection.

Indirect labeling involves the use of a protein, such as a labeled antibody, which specifically binds to the gene product. Such antibodies include but are not limited to

polyclonal, monoclonal, chimeric, single chain, Fab fragments and fragments produced by a Fab expression library.

In some instances, a preliminary selection is performed to verify that the host cells have been successfully transformed/transfected. Following the introduction of the vector, cells are allowed to grow in enriched media, and are then switched to selective media. The selectable marker confers resistance to the selective agent, and thus, only those cells that successfully express the introduced sequences survive in the selective media. Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in tk- or apr- cells, respectively (see e.g. Wigler, M. et al. (1977) "Transfer of purified herpes virus thymidine kinase gene to cultured mouse cells." *Cell* 11:223-32; Lowy, I. et al. (1980) "Isolation of transforming DNA: cloning the hamster aprt gene." *Cell* 22:817-23). Also antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate,; *neo* confers resistance to the aminoglycosides, neomycin and G-418, and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (see Wigler, M. et al. (1980) "Transformation of mammalian cells with an amplifiable dominant-acting gene." *PNAS* 77:3567-70; Colbere-Garapin, F. et al (1981) "A new dominant hybrid selective marker for higher eukaryotic cells." *J. Mol. Biol.* 150:1-14). Additional selectable genes have been described, e.g. *trpB* and *hisD*, which alter cellular requirements for metabolites. Visible markers, e.g. anthocyanins, green, red or blue fluorescent proteins (Clontech), B glucuronidase and its substrate B glucuronide, or luciferase and its substrate luciferin, may also be used. Resistant clones containing stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Host cells transformed/transfected with nucleotide sequences encoding for the perturbagen or target of interest may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. For example, the protein produced by a transformed transfected cell may be secreted when the selected expression vector incorporates signal sequences that direct secretion of the perturbagen through a prokaryotic or eukaryotic cell membrane.

Signal sequences also may be selected so as to direct the perturbagen to a particular intra-cellular compartment (Bradshaw, R.A. (1989) "Protein translocation and turnover in eukaryotic cells." *Trends Biochem Sci* 14(7):276-9). Perturbagen sequences may be isolated or purified from recombinant cell culture by methods heretofore employed for other proteins, e.g. native or reducing SDS gel electrophoresis, salt precipitation, isoelectric focusing, immobilized pH gradient electrophoresis, solvent fractionation, and chromatography such as ion exchange, gel filtration, immunoaffinity, and ligand affinity.

H. Host Cells

Host cell lines for use in the methodology described herein typically embody desirable traits such as 1) short cell cycle (i.e. 20-36 hr. doubling time), 2) amenability to high throughput procedures (e.g. FACS) without undue loss of membrane integrity or viability, 3) susceptibility to standard techniques designed to introduce reporter constructs and other forms of foreign DNA, and 4) exhibition of a readily selected phenotype (or its correlative marker gene expression). In this example, it is also advantageous if the host cell line chosen for perturbagen selection contains a physiology that is neither dependent upon, nor inordinately sensitive to, alterations in the β-catenin - TCF4 pathway. For example, if a cell responds to one or more of the components of the bioassay (e.g. β-catenin S45Y) or a perturbagen acting on said pathway of interest, by exhibiting a lower viability and/or capacity to compete in the population, for example, reducing said cell's growth rate, such a cell line would be less desirable than others that were insensitive to such changes. One non-limiting example of a satisfactory and acceptable cell line is HEK 293. HEK 293 is highly susceptible to retroviral infection and other methods of introducing foreign genetic materials and can express/maintain said materials for long periods of time using a variety of selectable markers common to the field (e.g. neomycin, puromycin). In addition, previous studies that elucidated some of the key functions of β-catenin were performed in HEK 293 cells, thus suggesting that the pathway can be altered in HEK293 cells without undue effects on cell viability (see, for example, Van Gassen, G. et al. (2000) Evidence that the beta catenin nuclear translocation assay allows for measuring presenilin 1 dysfunction." Mol Med 6(7): 570-80).

Reporter cell lines consist of a host cell that contains two components: i) a reporter gene operably-linked to a cis-acting promoter and ii) constitutive expression of a molecule that activates the TCF-β-catenin pathway. The terms "operably-associated" and "operably-linked" refer to functionally related nucleic acids. A promoter is "operably-associated" or "operably-linked" with a coding sequence if the promoter controls or regulates the transcription of the gene to which it is linked.

Favorable reporter lines exhibit several properties including 1) a large (>50-fold) "induced signal" to "uninduced signal" ratio, 2) a minimal number (<5%) of uninduced ("dim") cells when the population has been fully induced and, 3) a minimal number (<5%) of induced ("bright") cells when the populations is grown under non-inducing conditions. One example of an acceptable reporter cell line is a subline of HEK 293 referred to herein as S4535 (also referred to as S45-35). In addition to containing an activated form of β-catenin (e.g. β-cat S45Y), clone S4535, contains an artificial or synthetic promoter made up of four tandem repeats of the TBE-2 sequence (5' GCTTTGATC), operably-linked in cis to the reporter gene, Green Fluorescent Protein (EGFP, Clontech, also, see, He, T.C. et al. (1998) "Identification of c-Myc as a target of the APC pathway." *Science*, 281: 1509-12). Though this arrangement of promoter sequences and reporter gene performs adequately in the bioassay described, alternative components can be substituted to create equally efficient reporter lines. For instance, the regulatory region of the reporter need not be limited to the TBE-2 element as variations in both the sequence and number of tandemly-aligned TBE-2 or TBE-2-related cassettes can be successfully employed to create alternative β-catenin-TCF reporters. Furthermore, naturally occurring promoters such as those that lie upstream of the coding sequences of c-Myc, MMP-7, Cyclin D1 and other genes that respond to activation of the β-catenin-TCF pathway, can be fused to a variety of reporter genes (e.g. GFP, YFP, BFP, luciferase, β-galactosidase) to create functional reporter constructs (see, for example, Brabletz, T. et al. (1999) "beta-catenin regulates the expression of the matrix metalloproteinase-7 in human colorectal cancer." *Am J Pathol* 155(4):1033-8). Thus, the size of the β-catenin-TCF-responsive fragment can vary considerably. Synthetic promoters, which often consist of variations of the β-catenin-TCF consensus binding sequence can be relatively small (e.g. 17-70 nucleotides) while naturally occurring β-

catenin-TCF promoters sequences can be considerably larger (e.g. >500 bp) depending upon the breakpoints used to isolate the element.

The second component contained in the S4535 cell line is an activated form of β-catenin called β-cat S45Y. In the β-cat S45Y allele, standard molecular techniques have been employed to substitute a tyrosine residue for serine at amino acid residue 45. As a result of this alteration, a phosphorylation site that is normally critical for regulation of free, intracellular β-catenin levels is removed, thus creating a molecule that has an extended half-life and is, consequently, more active. As the change in β-cat S45Y does not alter the ability of the molecule to interact with TCF-4, addition of this mutant allele to HEK 293 cells leads to heightened expression of β-catenin-TCF-4 responsive genes. Thus, cells that have successfully incorporated both the β-cat S45Y expression vector and the TBE-2 reporter construct constitutively express GFP and can be selected by FACS. It should be noted that alternatives to the β-cat S45Y allele can be employed. Thus mutations that convert amino acid residues 33 or 37 (both serine residues) to tyrosine or phenylalanine respectively, can be utilized to activate the pathway in question.

Not all of the HEK293 cells that contain both the β-cat S45Y and the TBE-2 reporter construct respond equally to activation or deactivation of the β-catenin-TCF signal pathway. In some instances, the reporter construct may exhibit a constitutive phenotype due to insertion of the retroviral vector containing the reporter into a chromosomal position that contains a strong, constitutive, enhancer or promoter in a nearby region.

Other cells may exhibit little or no signal due to either i) insertion of the reporter element into a region of the genome that is transcriptionally silent or ii) introduction of a deleterious mutation in the β-cat S45Y during the process of transduction (the retroviral infection step is referred to herein as a transduction.) To eliminate these non-responsive cells and identify a clonal line that is readily modulated by perturbagens, a simple procedure is undertaken (Figure 3). In the first step, HEK 293 cells containing the reporter construct and the activated form of β-catenin undergo multiple cycles of FACS to isolate GFP^+ (bright) cells. In this fashion, it is assured that β-cat S45Y is intact and that the reporter construct has not been inserted into a region of the genome that is

transcriptionally silent. In the second step, individual clones that express high levels of GFP are obtained by plating cells at low density. After expanding these clones to a sufficient population, small samples of each are then transduced with a third retroviral construct containing a dominant negative allele of TCF4 called TCF4Δ30. The TCF4Δ30 cDNA is a truncated form of TCF4 in which 30 amino acids responsible for the interaction with β-catenin have been removed from the N-terminus of the molecule. As a result of this deletion, TCF4Δ30 is capable of binding to the TCF4 DNA consensus binding site (e.g. the TBE2 sequence of the reporter) yet is unable to induce transcription due to the absence of an activation domain normally provided by β-catenin. Samples of each clone containing all three constructs are then analyzed by FACS to determine which of the clones are responsive to the dominant negative effects of TCF4Δ30. Cultures that respond to the presence of TCF4Δ30 by shifting the fluorescent intensity of the population from "bright" to "dim", represent clones in which both the reporter and the β-cat S45Y allele of β-catenin have been inserted into chromosomal positions that permit modulation of the signal pathway. By returning to the original culture (e.g. +β-cat S45Y, + TBE2-reporter, -TCF4Δ30) from which these clones were derived, suitable reporter cell populations (e.g. S4535) can be expanded and utilized for subsequent perturbagen screens.

As one familiar with the art is aware, there are several variations to the procedures described above that could be used to achieve the same end results. For instance, FACS could be replaced with antibody affinity chromatography methods (using cell surface localized reporter) to segregate responsive from non-responsive clones (see, for example, Larsson, P.H. et al. (1989) "Improved cell depletion in a panning technique using covalent binding of immunoglobulins to surface modified polystyrene dishes." *J Immunol Methods*. 116(2):293-8; Contractor, S.F. et al. (1988) "Human placental cells in culture: a panning technique using a trophoblast-specific monoclonal antibody for cell separation." *J Dev Physiol.*10(1):47-51). Alternatively, the gates used to sort bright (or dim) cells in the FACS procedures could be adjusted to increase or decrease the enrichment procedures and thus alter the population of cells being collected and studied (see, for example Shapiro, H.M. (1995) "Practical Flow Cytometry" Wilely-Liss publishers). Furthermore, alleles other than β-catenin S45Y and TCFΔ30.

In addition, it is understood that there are many other suitable host cell lines including, but not limited to, transformed and/or immortalized cell lines derived from (i) HeLa (ATCC # CCL-2) and (ii) the CHO (ATCC # CCL-61, see, for example Tetsu et al. (1999) "β-catenin regulates expression of cyclinD1 in colon carcinoma cells ." *Nature* 398 (6726) 422-426; Sadot et al. (1998) , "Inhibition of β-catenin-mediated transactivation by cadherin derivatives." *PNAS* 95, 15339-15344). Any cell line such as these can be substituted as a host cell in the invention and can readily be screened to identify β-catenin-TCF reporter sublines. In addition, secondary cell lines may be used to study the effects of perturbagens. Thus, for instance, perturbagens isolated in HEK 293 cells may be introduced into, for example, HT-29 or SW620 colorectal adenocarcinoma cells (ATCC # HTB-38 and CCL-227 respectively) to study the effects of the perturbagen on cell lines that have a well-documented dependence on the β-catenin-TCF-APC pathway and/or have natural mutations in said pathway.

Lastly, critical components needed to construct a reporter cell line can be introduced into the cell by a variety of methodologies. For instance, episomal vectors carrying each of the components can be transformed into the cell type of choice and selected to identify rare events in which the vector becomes stably integrated into the genome. More preferably, a population of host cells containing the relevant constructs can be constructed using standard retroviral technology (Palu, G. et al. (2000) "Progress with retroviral gene vectors." *Rev Med Virol.* 10(3):185-202).

I. Screening for Biological Activity

The phenotypic assay described herein selects for perturbagens that modulate the β-catenin-TCF pathway. The procedures used to screen libraries for such perturbagens include: 1) introducing perturbagen encoding sequences (libraries) into clonal reporter cell lines containing both the activated form of β-catenin and the TBE-2 regulated reporter constructs; 2) growing said cells under the appropriate conditions necessary to identify perturbagens that repress expression of the reporter; 3) screening said cells by FACS or alternative high-throughput methods in order to segregate cells with the appropriate phenotype (e.g. "dim"); 4) re-isolating perturbagen encoding sequences from sorted cell populations by various techniques (e.g. PCR) and constructing new, retroviral sublibraries from the PCR product; 5) enriching for perturbagens by recycling said

sequences through the screen; and optionally 6) performing secondary assays to test specificity and scope of the agent.

Various methods and instrumentation familiar to those who are skilled in the art are used to screen and test perturbagens. The media, supplements, and reagents used in culturing, packaging, and maintenance of HEK 293 cells, HS293gp packaging cell lines, and additional lines (e.g. HT29, SW620) can be purchased from a variety of commercial sources (Life Technologies, Clonetics, Cocalico Biologicals Inc., ATCC). It should be noted that although a particular set of procedures and media formulations are used in the work described herein, alternatives can be substituted with little or no effect. For instance, in most cases, retroviral packaging was accomplished using $CaCl_2$. Though this is the preferred method of introducing retroviral vectors into 293gp packaging cells, alternative procedures such as LipofectAmine, may be used. Molecular techniques used in procedures such as genomic DNA isolation, PCR amplification, DNA endonculease digestion, ligation, cloning, and sequencing utilize common reagents that are supplied commercially (see, for example, Qiagen, New England BioLabs, Stratagene). Fluorescent activated cell sorting and analysis may be performed on a Coulter EPICS Elite Cell Sorter using EXPO software. Again, alternative reagents and equipment, such as the MoFlo^R High-Speed Cell Sorter (Cytomation), are compatible with these procedures and may be substituted with little or no effect.

To identify agents that deactivate the pathway, a retroviral expression library is introduced into Clone S4535 cells and grown/expanded over the course of several days in a selective environment (Figure 4). Under these conditions the vast majority of cells (>99%) do not contain a relevant perturbagen molecule and are "bright" due to the expression of high levels of the correlative reporter gene (e.g. red, green, or blue fluorescent protein). In contrast, a small fraction of the population appear "dim", a phenotype that results from, for instance, i) the presence of a perturbagen that inhibits or disrupts the β-catenin-TCF4-APC pathway or ii) loss of reporter expression due to cell death. "Dim" cells are then separated from the rest of the population by FACS and processed to re-isolate the perturbagen encoding sequences. Subsequently, these sequences are recycled through the bio-assay to further enrich for perturbagens that disrupt the pathway.

Several methods may be used to retrieve the perturbagen sequences from cells that have been sorted. For instance, perturbagen-encoding sequences may be recovered by PCR (see, for example, Schott, B. (1997) "Efficient recovery and regeneration of integrated retroviruses." *Nucleic Acids Res.* 25(14):2940-2). To accomplish this, genomic DNA (derived from cells taken from the FACS sorting procedures) is used as the template for PCR amplification. Complex mixtures with diversities of greater than 50,000 can be amplified efficiently using oligonucleotide primers that flank the perturbagen encoding sequence. These sequences can subsequently be ligated into an appropriate retroviral vector, and introduced into a fresh population of, e.g., S4535 cells for additional rounds of screening and enrichment. Alternatively, retrieval of the perturbagen may be accomplished by reactivating the inserted retroviral vector that contains the perturbagen-encoding sequence. Specifically, host cells containing the perturbagen-encoding retrovirus are transformed with sequences that encode the necessary retroviral gag, pol and envelope proteins. As a result of these procedures, retroviral virions that contain the perturbagen-encoding sequences are released and can be isolated in the form of a viral supernatant. These supernatants can then be utilized to infect fresh populations of, e.g., S4535 cells to recycle the sequences through the screen for additional enrichment.

Secondary cell lines may optionally be employed to test individual perturbagens for the ability to down-regulate β-catenin-TCF targets. For instance, perturbagens isolated in HEK 293 cells can be introduced into HT29 or SW610 colon cancer cells carrying a TBE-2-GFP reporter construct, and studied to determine whether the perturbagen's action is confined to a single cell type (e.g. HEK 293) or is broader in its application. In addition, these same cell lines can be used to study whether the perturbagen induced down-regulation of the β-catenin-TCF-APC pathway stimulates secondary phenotypes (such as cell cycle arrest and/or apoptosis) in cells where the β-catenin-TCF-APC pathway plays a more critical role. As one non-limiting example of how this might be accomplished, perturbagen sequences that induce a shift in the reporter expression levels could be introduced into (for instance) HT29 adenocarcinoma cells and tested for cytostatic or cytotoxic properties in a high throughput assay. In still another alternative, the effects of perturbagens on various aspects of tumor progression including

cell migration in colonic crypts, frequency of polyp formation, metastasis and other aspects colon cancer development, can be studied in whole animals by introducing perturbagen encoding sequences into either wild type or mutant mice carrying defects in one or more genes involved in the β-catenin-APC-TCF pathway. (e.g transgenic mice, see, for example, Sarao R, and Dumont DJ. (1998) "Conditional transgene expression in endothelial cells." *Transgenic Res.* 7(6):421-7; Wight D.C., Wagner T.E. (1994) "Transgenic mice: a decade of progress in technology and research." *Mutat Res* 307(2):429-40; Edelmann W. et al. (1999) "Tumorigenesis in Mlh1 and Mlh1/Apc1638N mutant mice." *Cancer Res* 59(6):1301-7).

J. Cellular targets

In other embodiments, the invention encompasses the polypeptide, ribonucleotide, or polynucleotide sequence of the target (or fragment of each target) that is identified with each perturbagen agent, as well as the gene encoding each target and relevant fragments of said gene.

Targets of specific perturbagens may be identified by several means. For instance, perturbagens can be modified with homo- or hetero- bifunctional coupling reagents and targets can be identified by chemical cross-linking techniques (see, for example, Tzeng, M.C. et al. (1995) "Binding proteins on synaptic membranes for crotoxin and taipoxin, two phospholipases A2 with neurotoxicity." *Toxicon.* 33(4):451-7; Cochet, C. et al. (1988) "Demonstration of epidermal growth factor-induced receptor dimerization in living cells using a chemical covalent cross-linking agent." *J Biol Chem.* 263(7):3290-5). Alternatively, one may use various techniques in column affinity chromatography, immunoprecipitation, or one of several high throughput peptide array platforms, to isolate peptides that react with the target of choice (see, for example, Hentz, N.G. and Daunert, S. (1996) "Bifunctional fusion proteins of calmodulin and protein A as affinity ligands in protein purification and in the study of protein-protein interactions." *Anal Chem.* 68(22):3939-44; Figeys D and Pinto D. (2001) "Proteomics on a chip: promising developments." *Electrophoresis* 22(2):208-16' Bichsel V.E. et al. "Cancer proteomics: from biomarker discovery to signal pathway profiling." *Cancer* J 7(1):69-78). In some instances, a particular phenotype may be the result of a perturbagen differentially regulating a distinct combination of genes. For example, a perturbagen

might, through its interaction with a particular transcription factor which, in turn, recognizes a particular DNA promoter sequence, elevate the expression of two or more target genes that act in concert to elicit a unique phenotype (e.g. viral resistance). In these cases, each of the genes whose levels of expression are altered by the perturbagen can be considered to be perturbagen targets. Such targets can be identified by a variety of techniques including (but not limited to) SAGE and expression profiling via microarray analysis (see, for instance, Cummings C.A. and Relman D.A. (2000) "Using DNA Microarrays to Study Host-Microbe Interactions." *Emerg Infect Dis.* 6(5):513-525; Yamamoto M. et al. (2001) "Use of serial analysis of gene expression (SAGE) technology. *J Immunol Methods.* 2001 Apr;250(1-2):45-66).

A preferred method of target identification involves application of variants of the standard two-hybrid technology. See, e.g., U.S.S.N. 09/193,759 and WO 00/29565 "Methods for validating polypeptide targets that correlate to cellular phenotypes", the entire disclosures of which are incorporated by reference herein. Generally stated, the two-hybrid procedure is a quasi-genetic approach to detecting binding events. This assay often is performed in yeast cells (although it can be adapted for use in mammalian and bacterial cells), and relies upon constructing two vectors; the first having an interaction probe or bait (that in this case, will be the perturbagen) that typically is fused to a DNA binding domain ("BD") moiety, and a second vector having an interaction target or prey (a cDNA library) that is typically fused to a DNA transcriptional moiety (the "activation domain" or "AD"). Neither of the two fusion proteins can, individually, induce transcription of the reporter gene. Yet when the bait and prey interact, the AD and BD moieties are brought into sufficient physical proximity to result in transcription of a reporter gene (e.g., the *His3* gene or lacZ gene) located downstream of the bound complex (Figure 5). Prey/bait interactions are then detected by identifying yeast cells that are expressing the reporter gene – e.g. which express lacZ or are able to grow in the absence of histidine.

A variety of yeast host strains known in the art are suitable for use for identifying targets of individual perturbagens. One of ordinary skill will appreciate that a number of factors may be considered in selecting suitable host strains, including but not limited to (1) whether the host cells can be mated to cells of opposite mating type (i.e., they are

haploid), and (2) whether the host cells contain chromosomally integrated reporter constructs that can be used for selections or screens (e.g., *His3* and *LacZ*). Although mating can be desirable in some embodiments, it is not strictly necessary for purposes of practicing the present invention. For example, the mating procedures can be eliminated by introducing the bait and prey constructs into a single yeast cell, whereupon the screens can be performed on the haploid cell.

Generally, either *Gal4* strains or *LexA* host strains may be used with the appropriate reporter constructs. Representative examples include strains yVT 69, yVT 87, yVT96, yVT97, yVT98 and yVT99, yVT100, yVT360. Additionally, those of ordinary skill will appreciate that the host strains used in the present invention may be modified in other ways known to the art in order to optimize assay performance. For example, it may be desirable to modify the strains so that they contain alternative or additional reporter genes that respond to two-hybrid interactions.

The following host yeast strains are thus constructed to have the indicated characteristics:

YVT69: yVT69 (mat □, ura3-52, his3-200, ade2-101, trp1-901, leu2-3, 112, gal4Δ, met$^-$, gal80Δ, URA3::$GAL1_{UAS}$-$GAL1_{TATA}$-lacZ) was obtained from Clontech (Y187).

YVT87: yVT87 (Mat-α ura3-52,his3-200,trp1-901 ,$LexA_{op\ (x6)}$-LEU2-3,112) was obtained from Clontech (EGY48).

YVT96: The starting strain was YM4271 (Liu, J. et al., 1993) MATa, ura3-52 his3-200 ade2-101 ade5 lys2-801 leu2-3, 112 trp1-901 tyr1-501 gal4Δ gal80Δ ade5::hisG. YM4271 was converted to yVT96, MATa ura3-52 his3-200 ade 2-101 ade5 lys2::GAL2-URA3 leu2-3, 112 trp1-901 tyr1-501 gal4D gal80Δ ade5::hisG by homologous recombination of Reporter 1 to the LYS2 locus. The integration is confirmed by PCR.

YVT97: The starting strain is YM4271 (Liu, J. et al., 1993) MATa, ura3-52 his3-200 ade2-101 ade5 lys2-801 leu2-3, 112 trp1-901 tyr1-501 gal4Δ gal80Δ ade5::hisG. YM4271 will be converted to yVT97, MATα ura3-52 his3::GAL1 or GAL7-HIS3 ade2-101 ade5 lys2-801 leu2-3, 112 trp1-901 tyr1-501 gal4Δ gal80Δ ade5::hisG by the steps of (a) converting from MATa to MATα via transient expression of the HO endonuclease, *Methods in Enzymology* Vol. 194:132-146 (1991) and (b) integrating either of Reporters

3 or 4 at the HIS3 locus via homologous recombination. The integration is confirmed by PCR.

YVT98: The starting strain was EGY48 (Estojak, J. Et al., 1995) MATα, ura3 his3 trp1 leu2::LexAop(x6)-LEU2. EGY48 was converted to strain yVT98 MATα ura3 his3 trp1 leu2::lexAop(x6)-LEU2 lys2::lexAop(8x or 2x)-LacZ by homologous recombination of Reporter 6 into the LYS2 locus.

YVT99: The starting strain was EGY48 (Estojak, J. Et al., 1995) MATα, ura3 his3 trp1 leu2::LexAop(x6)-LEU2. EGY48 was converted to strain yVT99 MATa ura3 his3 trp1 leu2::lexAop(x6)-LEU2 lys2::lexAop(8x or 2x)-URA3 by homologous recombination of Reporter 2 into the LYS2 locus and by switching the mating type from MATα to MATa via transient expression of the HO endonuclease.

YVT100: The starting strain was YM4271 (Liu, J. et al., 1993) MATa, ura3-52 his3-200 ade2-101 ade5 lys2-801 leu2-3, 112 trp1-901 tyr1-501 gal4Δ gal80Δ ade5::hisG. YM4271 was converted to yVT100, MATa ura3-52 his3-200 ade2-101 ade5 lys2::lexAop(8x or 2x)-URA3 leu2-3, 112 trp1-901 tyr-501 gal4Δ gal80Δ ade5::hisG by homologous recombination of Reporter 2 to the LYS2 locus. The integration was confirmed by PCR.

YVT360: yVT360 (mat a, trp1-901, leu2-3,112, ura3-52, his3-200, gal4 Δ, gal 80Δ, LYS2::$GAL1_{UAS}$-$GAL1_{TATA}$-HIS3, $GAL2_{UAS}$-$GAL2_{TATA}$-ADE2, URA3:$MEL1_{UAS}$-$MEL1_{TATA}$-lacZ) was obtained from Clontech (AH109).

Exemplary yeast-reporter strains are constructed using a variety of standard techniques. Many of the starting yeast strains already carry multiple mutations that lead to an auxotrophic phenotype (e.g. ura3-52, ade2-101). When necessary, reporter constructs can be integrated into the genome of the appropriate strain by homologous recombination. Successful integration can be confirmed by PCR. Alternatively, reporters may be maintained in the cells episomally.

The yeast two-hybrid reporter gene typically is fused to an upstream promoter region that is recognized by the BD, and is selected to provide a marker that facilitates screening. Examples include the *lacZ* gene fused to the *Gal1* promoter region and the *His3* yeast gene fused to *Gal1* promoter region. A variety of yeast two-hybrid reporter constructs are suitable for use in the present invention. One of ordinary skill will

appreciate that a number of factors may be considered in selecting suitable reporters, including whether (1) the reporter construct provides a rigorous selection (i.e., yeast cells die in the absence of a protein-protein or peptide-protein interaction between the bait and prey sequences), and/or (2) the reporter construct provides a convenient screen (e.g., the cells turn color when they harbor bait and prey sequences that interact). Examples of desirable reporters include (1) the *Ura*3 gene, which confers growth in the absence of uracil and death in the presence of 5-fluoroorotic acid (5-FOA); (2) the *His3* gene, which permits growth in the absence of histidine; (3) the *LacZ* gene, which is monitored by a colorimetric assay in the presence/absence of beta-galactosidase substrates (e.g. X-gal); (4) the *Leu2* gene, which confers growth in the absence of leucine; and (5) the *Lys2* gene, which confers growth in the absence of lysine or, in the alternative, death in the presence of α-aminoadipic acid. These reporter genes may be placed under the transcriptional control of any one of a number of suitable cis-regulatory elements, including for example the *Gal2* promoter, the *Gal1* promoter, the *Gal7* promoter, or the *LexA* operator sequences.

The following are exemplary, non-limiting examples of such reporter constructs.

Reporter 1 - (pVT85): This reporter comprises the URA3 gene under the transcriptional control of the yeast *Gal*2 upstream activating sequence (UAS). In order to facilitate integration of this reporter into the yeast chromosome in place of the *Lys*2 coding region, the *Gal*2-*Ura*3 construct is flanked on the 5' side by the 500 base pairs that lie immediately upstream of the coding region of the LYS2 gene and on the 3' side by the 500 base pairs that lie immediately 3' of the coding region of the LYS2 gene. The entire vector is also cloned into the yeast centromere containing vector pRS413 (Sikorski, RS and Hieter, P., *Genetics* 122(1):19-27 (1989) and can therefore be used episomally. This reporter is intended for use with a *Gal*4-based two-hybrid system, *e.g.*, Fields, S. and Song, O., *Nature* 340:245-246 (1989).

Reporter 2 - (pVT86): This reporter is identical to reporter #1 except that the GAL2 UAS sequences have been replaced with regulatory promoter sequences that contain eight *Lex*A operator sequences (Ebina et al., 1983). The number of *Lex*A operator sequences in this reporter may either be increased or decreased in order to obtain the

appreciate that a number of factors may be considered in selecting suitable reporters, including whether (1) the reporter construct provides a rigorous selection (i.e., yeast cells die in the absence of a protein-protein or peptide-protein interaction between the bait and prey sequences), and/or (2) the reporter construct provides a convenient screen (e.g., the cells turn color when they harbor bait and prey sequences that interact). Examples of desirable reporters include (1) the *Ura*3 gene, which confers growth in the absence of uracil and death in the presence of 5-fluoroorotic acid (5-FOA); (2) the *His3* gene, which permits growth in the absence of histidine; (3) the *LacZ* gene, which is monitored by a colorimetric assay in the presence/absence of beta-galactosidase substrates (e.g. X-gal); (4) the *Leu2* gene, which confers growth in the absence of leucine; and (5) the *Lys2* gene, which confers growth in the absence of lysine or, in the alternative, death in the presence of α-aminoadipic acid. These reporter genes may be placed under the transcriptional control of any one of a number of suitable cis-regulatory elements, including for example the *Gal2* promoter, the *Gal1* promoter, the *Gal7* promoter, or the *LexA* operator sequences.

The following are exemplary, non-limiting examples of such reporter constructs.

Reporter 1 - (pVT85): This reporter comprises the URA3 gene under the transcriptional control of the yeast *Gal*2 upstream activating sequence (UAS). In order to facilitate integration of this reporter into the yeast chromosome in place of the *Lys*2 coding region, the *Gal2-Ura*3 construct is flanked on the 5' side by the 500 base pairs that lie immediately upstream of the coding region of the LYS2 gene and on the 3' side by the 500 base pairs that lie immediately 3' of the coding region of the LYS2 gene. The entire vector is also cloned into the yeast centromere containing vector pRS413 (Sikorski, RS and Hieter, P., *Genetics* 122(1):19-27 (1989) and can therefore be used episomally. This reporter is intended for use with a *Gal*4-based two-hybrid system, *e.g.*, Fields, S. and Song, O., *Nature* 340:245-246 (1989).

Reporter 2 - (pVT86): This reporter is identical to reporter #1 except that the GAL2 UAS sequences have been replaced with regulatory promoter sequences that contain eight *Lex*A operator sequences (Ebina et al., 1983). The number of *Lex*A operator sequences in this reporter may either be increased or decreased in order to obtain the

optimal level of transcriptional regulation. This reporter is intended to be used within the general confines of the LexA-based interaction trap devised by Brent and Ptashne.

Reporter 3 - (pVT87): This reporter is comprised of the yeast *His3* gene under the transcriptional control of the yeast *Gal1* upstream activating sequence (UAS). In order to facilitate integration of this reporter into the yeast chromosome in place of the *His3* coding region the *Gal1-His3* construct is flanked on the 5' side by the 500 base pairs (bp) immediately upstream of the *His3* coding region and on the 3' side by the 500 bp immediately 3' of the *His3* coding region. The entire reporter is also cloned into the yeast centromere containing vector pRS415 and can therefore be used episomally. This reporter is intended for use with a *Gal4*-based two-hybrid system.

Reporter 4 - (pVT88): This reporter is identical to Reporter 3 except that the *His3* gene is under the transcriptional control of *Gal7* UAS sequences rather than the *Gal*1 UAS. The reporter is used with a *Gal4*-based two-hybrid system.

Reporter 5 - (pVT89): This reporter contains the bacterial *LacZ* gene under the transcriptional control of the *Gal1* UAS. The entire reporter will be cloned into a yeast centromere-using vector, e.g., pRS413, and is used episomally.

Reporter 6 - (pVT90): This reporter consists of the *LacZ* gene under the transcriptional control of eight *LexA* operator sequences. As for Reporter 2, the number of *LexA* operator sequences in this reporter may either be increased or decreased in order to obtain optimal levels of transcriptional regulation. Two features of this reporter facilitate integration of the reporter into the yeast chromosome in place of the *Lys2* coding region. First, it is flanked on the 5' side by the 500 base pairs that lie immediately upstream of the coding region of the *Lys2* gene and on the 3' side by the 500 base pairs that lie immediately 3' of the coding region of the *Lys2* gene. Second, the neomycin (NEO) resistance gene has been inserted between the 5' *Lys2* sequences and the LexA promoter sequences. This reporter is used in conjunction with a *LexA*-based interaction trap, *e.g.*, Golemis, E.A., et al., (1996), "Interaction trap/two hybrid system to identify interacting proteins." *Current Protocols in Molecular Biology*, Ausebel et al., eds., New York, John Wiley & Sons, Chap. 20.1.1-20.1.28.

In other embodiments, perturbagen-induced phenotypes may be the result of RNA-RNA, RNA-polypeptide, polypeptide-DNA, or RNA-DNA interactions. In cases

such as these, variations of the original two-hybrid theme may be applied to identify the target of the phenotypic probe. (See, for example, Li, J.J. and Herskowitz, I. (1993) Isolation of Orc6, a Component of the Yeast Origin Recognition Complex by a One-Hybrid System. *Science*, 262:1870-1874; Svinarchuk, F. et al. (1997) "Recruitent of transcription factors to the target site by triplex-forming oligonucleotides." *NAR* 25: 3459-3464; Segupta, D.J. et al. (1999) "Identification of RNAs that bind to a specific protein using the yeast three-hybrid system." *RNA* 5:596-601; Harada, K. et al. (1996) "Selection of RNA-binding peptides in vivo." *Nature* 14;380(6570):175-9; SenGupta, D.J. et al. (1996) "A three-hybrid system to detect RNA protein interactions in vivo." *PNAS* 93:8496-8501). For instance, if evidence exists that a perturbagen is acting as an anti-sense agent, it is necessary to construct a system where the association of the DNA binding domains and the transcriptional activation domains is dependent upon and RNA-RNA interaction. To accomplish such a screen, four unique vectors are created (Figure 6). The first vector consists of the DNABP (e.g. GAL4 BD) described previously, linked to a specific RNA binding protein, arbitrarily called "RNABP-A" (e.g. the Rev responsive element RNA binding protein, RevM10, see Putz, U. et al. (1996) "A tri-hybrid system for the analysis and detection of RNA-protein interactions." *NAR* 24:4838-4840). Vector #2 contains the transcriptional activation domain (e.g. GAL4 AD) linked to a second RNA binding protein ("RNABP-B", e.g. the MS2 coat protein of the MS2 bacteriophage, see for example, SenGupta, D.J. et al. (1996) "A three hybrid system to detect RNA-protein interactions in vivo." *PNAS* 93:8496-8501). The third vector encodes an RNA molecule that is recognized by RNABP-A (e.g. the RRE sequence, Zapp, M.L. and Green M.R/ "Sequence-specific RNA binding by the HIV-1 Rev protein (1989) *Nature*, 32:714-716) fused to a sequence encoding the RNA perturbagen, while the final vector encodes a fourth hybrid, the RNA sequence recognized by RNABP-B (e.g. the 21 base nucleotide RNA stem-loop structure of MS2, see Uhlenbeck, O.C. et. al. (1983) "Interaction of R17 coat protein with its RNA binding site for translational repression." *J. Biomol Struct. Dyn.* 1, 539-552) linked to a library of expressed sequences (e.g. a library of mRNA molecules). When all four vectors are stably maintained in a yeast cell containing the necessary reporter construct(s) (e.g. P_{GAL4}-LACZ), the cellular target RNA molecule of any given RNA perturbagen can be identified.

Target sequences or fragments thereof can vary greatly in size. Some target fragments can be as small as ten amino acids in length. Alternatively, target sequences can be greater than 10 amino acids but less than thirty amino acids in length. Still other targets can be greater than thirty amino acids in length but shorter than 60 amino acids in length. Still other targets are cellular proteins or subunits or domains therein of more than 60 amino acids in length. Still other targets are cellular proteins or subunits or domains there of more than 60 amino acids in length. Still other targets are cellular proteins or subunits or domains there of more than 60 amino acids in length. In addition, for reasons described previously, the sequences encoding targets can vary greatly due to allelic variation, duplications and closely related gene family members. That said, the invention also encompasses variants of said targets. A preferred target variant is one which has at least about 80%, alternatively at least about 90%, and in another alternative at least about 95% amino acid sequence identity to the original target amino acid sequence and which contains at least one functional or structural characteristic of the original target.

K. Databases

The compositions, relations and phenotypic effects yielded by the methodology described herein may advantageously be placed into or stored in a variety of databases. As one example, a database may include information about one or more targets identified by the methods herein, including for example sequence information, motif information, structural information and/or homology information. The database may optionally contain such information regarding perturbagen agents, and may correlate the perturbagen information to corresponding target information. Further helpful database aspects may include information regarding, e.g., variants or fragments of the above. The database may also correlate the indexed compounds to, e.g., immunoprecipitation data, further yeast n-hybrid interaction data, genotypic data (e.g., identification of disrupted genes or gene variants), and with a variety phenotypic data. Such databases are preferably electronic, and may additionally be combined with a search tool so that the database is searchable.

L. Production of antibodies

An additional embodiment of the invention includes antibodies that recognize the perturbagen itself, cellular targets of the perturbagen, or one or more epitopes of the

foregoing. Such reagents may include, but are not limited to, polyclonal, monoclonal, humanized, chimeric, and single chain antibodies, Fab fragments, $F(ab')_2$ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope-binding fragements of any of the above. Antibodies directed against perturbagens or cellular targets may be useful for a variety of purposes including i) therapeutics, ii) diagnostic assays, iii) cytoimmunology, iv) target identification, and v) purification.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans and others may be immunized by injection with a perturbagen, target or any fragment thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, and surface-active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of animals immunized with an antigen, such as a given perturbagen, target, or an antigenic functional derivative thereof. For the production of polyclonal antibodies, host animals such as those described above, may be immunized by injection with gene product supplemented with adjuvants as also described above.

Monoclonal antibodies that recognize perturbagens may be prepared using any technique that provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV hybridoma technique. (see, for example, Kohler, G. et al. (1975) "Continuous cultures of fused cells secreting antibody of predefined specificity." *Nature* 256:495-497; Kozbor, D. et al (1985) "Specific immunoglobulin production and enhanced tumorigenicity following ascites growth of human hybridomas." J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) *PNAS* 80:2026-2030; and Cole, S.P. et al. (1984) "Generation of human monoclonal antibodies reactive with cellular antigens" *Mol. Cell Biol.* 62:109-120).

In addition, one may use techniques developed for the production of chimeric antibodies, such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity. See, e.g., Morrison, S.L. et al. (1984) "Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains." *PNAS* 81:6851-6855); Neuberger, M.S. et al. (1984) "Recombinant antibodies possessing novel effector functions." *Nature* 312:604-608; and Takeda, S. et al. (1985) "Construction of chimeric processed immunoglobulin genes containing mouse variable and human constant region sequences." *Nature* 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce perturbagen-specific antibodies (see, e.g. Burton, D.R. (1991) "A large array of human monoclonal antibodies to type 1 human immunodeficiency virus from combinatorial libraries of asymptomatic seropositive individuals." *PNAS* 88:10134-10137). Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (see, for example, Orlandi, R. et al. (1989) "Cloning immunoglobulin variable domains for expression by the polymerase chain reaction." *PNAS* 86:3833-3837; Winter, G. et al. (1991) "Man-made antibodies." *Nature* 349: 293-299).

Antibody fragments that contain specific binding sites for perturbagens may also be generated. For example, such fragments include, but are not limited to $F(ab')_2$ fragments produced by pepsin digesting of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monclonal Fab fragments with the desired specificity. (See, for example, Huse, W.D. et al. (1989) "Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda." *Science* 246:1275-1281).

M. Screening Assays

The agents of the invention can be used to screen for drugs or compounds (small molecules) that mimic, or modulate the activity or expression of said phenotypic probes. The present invention may be employed in a process for screening for agents such as

agonists, *i.e.* agents that bind to and activate a β-catenin/Tcf pathway target, or antagonists, *i.e.* inhibit the activity or interaction of an β-catenin/Tcf pathway target with an endogenous or exogenous ligand. Thus, polypeptides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures as known in the art. Any methods routinely used to identify and screen for agents that can modulate receptors may be used in accordance with the present invention.

Like the perturbagen itself, such small molecule compounds may be used to treat disorders characterized by insufficient or excessive production of a target which has decreased or aberrant activity compared to the wild type entity. Thus, the invention provides a method for identifying modulators, i.e. candidate or test compounds or agents (e.g. peptidomimetics, small molecules or other drugs) that bind to the agent or its target, and have a stimulatory or inhibitory effect on the pathway(s) affected by said agent. *In vitro* systems may be designed to identify compounds capable of binding, e.g., a β-catenin pathway target gene product. Such compounds may include, but are not limited to, peptides made of D-and/or L-configuration amino acids (in, for example, the form of random peptide libraries; (*see e.g.*, Lam, *et al.*, *Nature*, 354:82-4 (1991)), phosphopeptides (in, for example, the form of random or partially degenerate, directed phosphopeptide libraries; *see*, *e.g.*, Songyang, *et al.*, *Cell*, 72:767-78 (1993)), antibodies, and small organic or inorganic molecules. Compounds identified may be useful, for example, in modulating the activity of β-catenin pathway target gene proteins, preferably mutant proteins; elaborating the biological function of the β-catenin pathway target gene protein; or screening for compounds that disrupt normal β-catenin pathway target gene interactions or themselves disrupt such interactions.

In one embodiment, the invention provides libraries of test compounds. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries, spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the one-bead one-compound library method; and synthetic library methods using affinity chromatography selection. The biological library approach is exemplified by peptide libraries, while the other four approaches are

applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) "Application of combinatorial library methods in cancer research and drug discovery." *Anticancer Drug Des.* 12:145).

Methods for the synthesis of molecular libraries can be found in the art, for example, in (i) De Witt, S.H. et al. (1993) "Diversomers: an approach to nonpeptide, nonoligomeric chemical diversity." *PNAS* 90:6909, (ii) Erb, E. et al. (1994) "Recursive deconvolution of combinatorial chemical libraries ."*PNAS* 91:11422, (iii) Zuckermann, R.N. et al. (1994) "Discovery of nanomolar ligands for 7-transmembrane G-protein-coupled receptors from a diverse N-(substituted)glycine peptoid library." *J. Med Chem.* 37: 2678 and (iv) Cho, C.Y. et al. (1993) "An unnatural biopolymer." *Science* 261:1303. Libraries of compounds may be presented in i) solution (e.g. Houghten, R.A. (1992) "The use of synthetic peptide combinatorial libraries for the identification of bioactive peptides." *BioTechniques* 13:412) ii) on beads (Lam, K.S. (1991) "A new type of synthetic peptide library for identifying ligand-binding activity." *Nature* 354:82), iii) chips (Fodor, S.P. (1993) "Multiplexed biochemical assays with biological chips." *Nature* 364:555), iv) bacteria (U.S. Patent # 5,223,409), v) spores (Patent Nos 5,571,698, 5,403,484, and 5,223,409), vi) plasmids (Cull, M.G. et al. (1992) "Screening for receptor ligands using large libraries of peptides linked to the C terminus of the lac repressor." *PNAS* 89:1865) or vii) phage (Scott, J.K. and Smith, G.P. (1990) "Searching for peptide ligands with an epitope library." *Science* 249: 386)

There are several methods for identifying small molecule compounds that mimic the action of the phenotypic probes. In one approach, an assay may be devised to directly identify agents that bind to, e.g., a βcatenin/TCF pathway target protein. Such direct binding assays generally involve preparing a reaction mixture of the βcatenin/TCF pathway target protein and the test compound under conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex that can be removed and/or detected in the reaction mixture. These assays can be conducted in a variety of ways. For example, one method to conduct such an assay would involve anchoring the βcatenin/TCF pathway target protein or the test substance onto a solid phase and detecting target protein/test substance complexes anchored on the solid phase at the end of the reaction. In one embodiment of such a method, the βcatenin/TCF

pathway target protein may be anchored onto a solid surface, and the test compound, which is not anchored, may be labeled, either directly or indirectly.

In practice, microtitre plates are conveniently utilized. The anchored component may be immobilized by non-covalent or covalent attachments. Non-covalent attachment may be accomplished simply by coating the solid surface with a solution of the protein and drying. Alternatively, an immobilized antibody, preferably a monoclonal antibody, specific for the protein may be used to anchor the protein to the solid surface. The surfaces may be prepared in advance and stored.

In order to conduct the assay, the nonimmobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (*e.g.*, by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously nonimmobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously nonimmobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; *e.g.*, using a labeled antibody specific for the previously nonimmobilized component (the antibody, in turn, may be directly labeled or indirectly labeled with a labeled anti-Ig antibody).

Alternatively, a reaction can be conducted in a liquid phase, the reaction products separated from unreacted components, and complexes detected; *e.g.*, using an immobilized antibody specific for a βcatenin/TCF pathway gene product or the test compound to anchor any complexes formed in solution, and a labeled antibody specific for the other component of the possible complex to detect anchored complexes. Compounds that are shown to bind to a particular βcatenin/TCF pathway gene product through one of the methods described above can be further tested for their ability to elicit a biochemical response from the a βcatenin/TCF pathway gene protein. Agonists, antagonists and/or inhibitors of the expression product can be identified utilizing assays well known in the art.

In another approach, perturbagen/target pairs are used to identify small molecule mimetics in a displacement assay format. Such assays can be based upon a variety of

technologies including, but not limited to i) ELISAs (see, for example, Rice, J.W. et al. (1996) "Development of a high volume screen to identify inhibitors of endothelial cell activation." *Anal Biochem* 241(2):254-9), ii) scintillation proximity assays (see, for example, Lerner, C.G. and Saiki, A.Y.C. (1996) "Scintillation proximity assay for human DNA topoisomerase I using recombinant biotinyl-fusion protein produced in baculovirus-infected insect cells." *Anal Biochem* 240(2):185-96), or iii) time-resolved fluorescence resonance energy transfer-based technology (see, for example, Fernandes, P.B. (1998) "Technological advances in high-throughput screening." *Curr Opin Chem Biol* 2(5):597-603; Hemmilä, "Time-resolved fluorometry - advantages and potentials in high throughput screening assays." "*High Throughput Screening*", J. Devlin (ed.). Marcel Dekker Inc, New York, pp. 361-76 (1997)). Two non-limiting examples of such assays, one homogeneous, LANCE™ (Stenroos, K. et al. (1997) "Homogeneous time resolved fluorescence energy transfer assay (LANCE) for the determination of IL-2-IL-2 receptor interaction." Abstract of Papers Presented at the 3rd Annual Conference of the Society for Biomolecular Screening, Sep., California), and one heterogeneous, DELFIA™ (MacGregor, I. et al. (1999) "Application of a time-resolved fluoroimmunoassay for the analysis of normal prion protein in human blood and its components." *Vox Sang* 77(2):88-96; Jensen, P.E. et al. (1998) "A europium fluoroimmunoassay for measuring peptide binding to MHC class I molecules." *J. Immunol. Methods* 215: 71-80; Takeuchi, T. et al. (1995) "Nonisotopic receptor assay for benzodiazepine drugs using time-resolved fluorometry." *Anal. Chem.* 67: 2655-8) are described as follows.

1. Lance™: Homogeneous Assay

To identify small molecules capable of disrupting the interaction between the perturbagen and its target, assays are designed to utilize the LANCE™ technology (commercially available from E. G. & G. Wallac.). LANCE™ is a homogeneous assay that is performed in solution and requires no wash steps to separate bound and unbound label. Briefly, the target is produced in large quantities and labeled with a lanthanide chelate (i.e. a fluorescent donor such as a Europium, (Eu) or Terbium (Tb) chelate). Concomitantly, the perturbagen is labeled with one of several fluorescent "acceptor" moieties that can be excited by the emissions of the donor molecule (e.g. allophycocyanin (APC) or rhodamine Rh, respectively). Most preferably, 1) the modification of either the

perturbagen or the target is not detrimental to the interaction between the two interacting molecules being studied and 2) the distance separating the donor and acceptor moieties when the perturbagen and the target are associated, is sufficiently close to permit FRET (typically 30-100 Angstroms). As an alternative to direct labeling of the perturbagen, monoclonal antibodies directed against the perturbagen can be labeled with Eu, thus allowing small molecule displacement assays to take place via indirect labeling procedures.

To identify small molecules capable of disrupting the interaction between the perturbagen and its target, the two labeled components are alliquoted into wells (1536 well format) at previously set, optimized conditions that will ensure 50% binding (Figure 7). Subsequently, each well is then exposed to one or more members of a large chemical combinatorial library and time-resolved measurements are taken using a Wallac 1420 Victor multilabel counter or equivalent fluoremeter. In wells that contain a small molecule that interferes with the interaction between the perturbagen and its target, the distance separating the donor and acceptor molecules is increased. As a result of this dissociation or displacement, the ability of the Eu emissions to excite the acceptor is compromised and the total fluorescence emitted by the acceptor is decreased.

2. DELFIA™: Heterogeneous Assay

Several variations of a heterogeneous assay (DELFIA™) using an immobilized substrate can be used as an alternative to LANCE™. In one non-limiting example, the target is immobilized to a solid support using a monoclonal antibody that has been labeled with Eu (Figure 8). Subsequent addition and binding of a rhodamine labeled perturbagen in the presence or absence of a candidate small organic displacement molecule is followed by several wash steps to remove unbound material. TR-FRET is then performed by exciting Eu and measuring the levels of Rh emissions. As an alternative to this procedure, the target is immobilized to the solid support using an unlabeled monoclonal antibody. Subsequently, an Eu-labeled perturbagen (+/- a candidate small organic displacement molecule) is added to each well and allowed to equilibrate, followed by a washing procedure to eliminate unbound Eu-labeled material. Once the well has been cleared of all unbound material, the bound Eu-perturbagen molecules are released and excited in the presence of commercially available

enhancement solutions (DELFIA™ Enhancement Solutions, Wallac). By comparing the levels of emissions in wells that contain members of the molecule library with standardized controls, small molecules that disrupt the interaction between the perturbagen and its target are identified.

Another preferred method for identifying small molecule mimetics makes use of a variation of the two-hybrid technology. As one non-limiting example of how a two-hybrid chemical screen is performed, the yeast host cells containing i) AD-perturbagen, ii) the BD-target, and iii) a reporter construct made up of a promoter recognized by the BD, functionally linked to, for instance, the gene encoding lacZ or ZsGreen (Clontech), are grown in liquid culture media and subjected to the test chemical. Assay plates are then incubated at 30°C for 48 hours and samples are scored by looking the expression of the marker by FACS or other conventional techniques. As an alternative, compounds that are attached to a solid support (e.g. beads) can be tested for their ability to rescue the growth phenotype in solution-based assays. Specifically, yeast cells modified for reverse genetic studies can be arrayed in nanodroplets (100-200 nanoliter volumes) that contain i) the selective elements of the medium (e.g. 5-FOA, cyclohexamide) and ii) one or more beads linked to a chemical library member. Subsequent photolysis of the chemical agent from the bead allows diffusion of the test molecules into the yeast cell and disruption of the two-hybrid interaction (see, Borchardt A et al. (1997) "Small molecule-dependent genetic selection in stochastic nanodroplets as a means of detecting protein-ligand interactions on a large scale." *Chemical Biology* 4(12):961-8; You, A.J. et al. (1997) "A miniaturized arrayed assay format for detecting small molecule-protein interactions in cells." *Chem Biol.* 4(12):969-75; Huang, J. Schreiber, S.L. (1997) "A yeast genetic system for selecting small molecule inhibitors of protein-protein interactions in nanodroplets." *PNAS* 94:13396-13401; Young, K. et al. (1998) : Identification of a calcium channel modulator using a high throughgput yeast two-hybrid screen." *Nature Biotechnology* 16:946-950).

L. Therapeutic Uses

Natural and synthetic chemotherapeutic derivatives have proven valuable in the treatment of a variety of forms of disease. For that reason, in one embodiment, perturbagens, fragments or derivatives of a perturbagen, small molecule mimetics of a

perturbagen, sequences encoding perturbagens, sequences that can hybridize to perturbagen encoding sequences, targets of the perturbagen, or agents that bind said target (e.g. antibodies) or portions thereof, may be utilized to treat or prevent a disorder that has previously shown sensitivity to treatment with chemotherapeutics and/or radiation therapy. Thus, for example, polypeptides or RNA molecules described herein can be used i) modulate cellular proliferation, ii) modulate cellular differentiation, iii) induce or modulate necrotic or apototic processes, or iv) sensitize cells to secondary compounds that induce either i), ii), or iii) by direct application of said agent. Examples of such disorders that may be aided by such agents include, but are not limited to cancers of the I) ovary, ii) liver, iii) endometrium, iv) colon and/or rectum, v) prostrate, vi) uterus, vii) esophagus, viii) kidney, ix) thyroid, x) stomach, xi) brain, and xii) skin (e.g. melanoma). In addition, agents identified in previously described screens may be applicable in treating a variety of other diseases that directly or indirectly involve components of the β-catenin /TCF/APC pathway. Thus, any of the agents of the invention may be administered to a subject to treat or prevent, for instance, Alzheimer's disease, CHRPE and related afflictions.

Ailments such as those described previously can be treated with the perturbagen or target directly, for example by administering a therapeutically effective dose of a proteinaceous agent intravenously or by other peptide delivery techniques known to the art. A therapeutically effective dose of a pharmaceutical composition comprising a substantially purified perturbagen, or a fragment thereof, or a small molecule mimetic, optionally in conjunction with a suitable pharmaceutical carrier, may be administered to a subject to treat or prevent a disorder previously shown to be related to the βcatenin/Tcf pathway. A "therapeutically effective" dose refers to that amount of the compound sufficient to result in amelioration of symptoms of the disease. A "pharmaceutical carrier" includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50}. Compounds that exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (*i.e.*, the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

Pharmaceutical compositions of the invention are formulated to be compatible with intended routes of delivery. Examples of routes of administration include parenteral e.g. intravenous, intradermal, subcutaneous, oral, inhalation, transdermal, topical, transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent, such as water for injection, saline solution, fixed oils, polyethylene, glycols, glycerine, propylene glycol, or other synthetic solvents, antibacterial agents such as benzyl alcohol or methyl parabens, antioxidants such as ascorbic acid or sodium bisulfite, chelating agents such as ethylenediaminetetraacetic acid, buffers such as acetates,

citrates, or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose.

Pharmaceutical compositions suitable for injectable use include aqueous solutions (where water-soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water Cremophor EL™ (BASF; Parsippany, NJ) or phosphate buffered saline (PBS). In all cases the composition must be sterile and should be fluid to the extent that easy syringability exists. Oral compositions can also be prepared using any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth, or gelatin; an excipient such as starch or lactose, disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate, a glidant such as colloidal silicon dioxide, a sweetening agent such as sucrose or saccharin, or a flavoring agent such as peppermint or orange flavoring. For administration by inhalation, the compounds are delivered in the form of an aerosol spray from a pressurized container or dispenser that contains a suitable propellant. Systemic administration can also be by transmucosal or transdermal means. For these methods of administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art and include, for example, bile salts and fusidic acid derivatives. Transmucosal administration can also be accomplished through the use of nasal sprays and suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled microencapsulated delivery system. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to specific cell surface epitopes) can also be used as pharmaceutically acceptable carriers. These can be

prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

Alternatively, such therapeutics can be administered indirectly, for example by gene therapy utilizing a gene or RNA sequence encoding a perturbagen, pathway target, or variant or fragment of the foregoing. For example, a vector capable of expressing a perturbagen or target, or a fragment or derivative thereof, may be administered to a subject to treat or prevent a disease. Expression vectors including, but not limited to, those derived from retroviruses, adenoviruses, adeno-associated viruses, or herpes or vaccinia viruses or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population (see, for example, Carter, P.J. and Samulski, R.J. (2000) "Adeno-associated viral vectors as gene delivery vehicles." *Int J Mol Med.* 6(1):17-27; Palu, G. et al. (2000) "Progress with retroviral gene vectors." *Rev Med Virol.* 10(3):185-202; Wu, N. and Ataai, M.M. (2000) "Production of viral vectors for gene therapy applications." *Curr Opin Biotechnol.* 11(2):205-8). Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. Patent 5,328,470) or by stereotactic injection (see, for example, Chen, S.H. et al. (1994) "Gene therapy for brain tumors: regression of experimental gliomas by adenovirus-mediated gene transfer in vivo." *PNAS* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system. .

M. Antisense, Ribozyme and Antibody Therapeutics

Other agents that may be used as therapeutics include any pathway target genes, associated expression product and functional fragments thereof. Additionally, agents that reduce or inhibit mutant pathway target gene activity may be used to ameliorate disease symptoms. Such agents include antisense, ribozyme, and triple helix molecules. Techniques for the production and use of such molecules are well known to those of skill in the art.

Anti-sense RNA and DNA molecules act to directly block the translation of mRNA by hybridizing to targeted mRNA and preventing protein translation. With respect to antisense DNA, oligodeoxyribonucleotides derived from the translation initiation site, *e.g.*, between the -10 and +10 regions of the β-catenin pathway target gene nucleotide sequence of interest, are preferred.

Ribozymes are enzymatic RNA molecules capable of catalyzing cleavage of specific RNAs. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by an endonucleolytic cleavage. The composition of ribozyme molecules must include one or more sequences complementary to a β-catenin pathway target gene mRNA, and must include the well known catalytic sequence responsible for mRNA cleavage. For this sequence, see U.S. Patent No. 5,093,246, which is incorporated by reference herein in its entirety. As such within the scope of the invention are engineered hammerhead motif ribozyme molecules that specifically and efficiently catalyze endonucleolytic cleavage of RNA sequences encoding anaphylatoxin C3a receptor gene proteins.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the molecule of interest for ribozyme cleavage sites that include the following sequences, GUA, GUU and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the β-catenin pathway target gene containing the cleavage site may be evaluated for predicted structural features, such as secondary structure, that may render the oligonucleotide sequence unsuitable. The suitability of candidate sequences may also be evaluated by testing their accessibility to hybridization with complementary oligonucleotides, using ribonuclease protection assays.

Nucleic acid molecules to be used in triple helix formation for the inhibition of transcription should be single stranded and composed of deoxyribonucleotides. The base composition of these oligonucleotides must be designed to promote triple helix formation via Hoogsteen base pairing rules, which generally require sizeable stretches of either purines or pyrimidines to be present on one strand of a duplex. Nucleotide sequences may be pyrimidine-based, which will result in TAT and CGC triplets across the three associated strands of the resulting triple helix. The pyrimidine-rich molecules provide

base complementarity to a purine-rich region of a single strand of the duplex in a parallel orientation to that strand. In addition, nucleic acid molecules may be chosen that are purine-rich, for example, containing a stretch of G residues. These molecules will form a triple helix with a DNA duplex that is rich in GC pairs, in which the majority of the purine residues are located on a single strand of the targeted duplex, resulting in GGC triplets across the three strands in the triplex.

Alternatively, the potential sequences that can be targeted for triple helix formation may be increased by creating a so called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex. It is possible that the antisense, ribozyme, and/or triple helix molecules described herein may reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by both normal and mutant β-catenin pathway target gene alleles. In order to ensure that substantially normal levels of β-catenin pathway target gene activity are maintained, nucleic acid molecules that encode and express β-catenin pathway target gene polypeptides exhibiting normal activity may be introduced into cells that do not contain sequences susceptible to whatever antisense, ribozyme, or triple helix treatments are being utilized. Alternatively, it may be preferable to co-administer normal β-catenin pathway target gene protein into the cell or tissue in order to maintain the requisite level of cellular or tissue β-catenin pathway target gene activity.

Anti-sense RNA and DNA, ribozyme, and triple helix molecules of the invention may be prepared by any method known in the art for the synthesis of DNA and RNA molecules. These include techniques for chemically synthesizing oligodeoxyribonucleotides and oligoribonucleotides well known in the art such as for example solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding the antisense RNA molecule. Such DNA sequences may be incorporated into a wide variety of vectors that incorporate suitable RNA polymerase promoters such as the T7 or SP6 polymerase promoters. Alternatively, antisense cDNA constructs that synthesize

antisense RNA constitutively or inducibly, depending on the promoter used, can be introduced stably into cell lines.

Various well-known modifications to the DNA molecules may be introduced as a means of increasing intracellular stability and half-life. Possible modifications include but are not limited to the addition of flanking sequences of ribonucleotides or deoxyribonucleotides to the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the oligodeoxyribonucleotide backbone.

Antibodies that are both specific for β-catenin pathway target gene protein, and in particular, mutant gene protein, and interfere with its activity may be used to inhibit mutant β-catenin pathway target gene function. Such antibodies may be generated against the proteins themselves or against peptides corresponding to portions of the proteins using standard techniques known in the art and as also described herein. Such antibodies include but are not limited to polyclonal, monoclonal, Fab fragments, single chain antibodies, chimeric antibodies, etc.

In instances where a β-catenin pathway target gene protein is intracellular and whole antibodies are used, internalizing antibodies may be preferred. However, lipofectin liposomes may be used to deliver the antibody or a fragment of the Fab region that binds to the β-catenin pathway target gene epitope into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to the target or expanded target protein's binding domain is preferred. For example, peptides having an amino acid sequence corresponding to the domain of the variable region of the antibody that binds to the β-catenin pathway target gene protein may be used. Such peptides may be synthesized chemically or produced via recombinant DNA technology using methods well known in the art (*see*, *e.g.*, Creighton, Proteins: Structures and Molecular Principles (1984) W.H. Freeman, New York 1983, *supra*; and Sambrook, *et al.*, 1989, supra). Alternatively, single chain neutralizing antibodies that bind to intracellular β-catenin pathway target gene epitopes may also be administered. Such single chain antibodies may be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population by utilizing, for example, techniques

such as those described in Marasco, *et al.*, *Proc. Natl. Acad. Sci. USA*, 90:7889-93 (1993).

N. Diagnostic Uses

The polynucleotides, polypeptides, variants, targets and antibodies to any one of these molecules can, in addition to previously mentioned therapeutic applications, be used in one or more of the following methods: 1) detection assays (e.g. chromosomal mapping, tissue typing, forensic biology), and 2) predictive medicine (e.g. diagnostic or prognostic assays, pharmacogenomics and monitoring clinical trials). Thus, for example, agents may be used to detect a specific mRNA or gene (e.g. in a biological sample) for a genetic lesion. Similarly, agents described herein may be applied to the field of predictive medicine in which diagnostic assays or prognostic assays, pharmacogenomics, and monitoring clinical trials are used for predictive purposes to thereby treat an individual prophylactically.

Accordingly, one aspect of the present invention relates to diagnostic assays for determining expression of a polypeptide or nucleic acid of the invention and/or activity of said agent of the invention, in the context of a biological sample to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant expression or activity of a polypeptide or polynucleotide of the invention.

Alternatively, the invention provides methods for detecting expression of a nucleic acid or polypeptide of the invention or activity of a polypeptide or polynucleotide of the invention in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmoacogenomics allows for the selection of agents (e.g. drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (e.g. the genotype of the individual examined to determine the ability of the individual to respond to a particular agent). Still another aspect of the invention pertains to monitoring the influence of agents (e.g. drugs or other compounds) on the expression or activity of a polypeptide or polynucleotide of the invention in clinical trials.

1. Detection Assays

Portions or fragments of the polynucleotide sequences of the invention can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to i) map their respective genes on a chromosome and, thus, locate gene regions associated with genetic diseases; ii) identify an individual from a minute biological sample (tissue typing); and iii) aid in forensic identification of biological samples.

a. Gene and Chromosome Mapping.

Once the sequence (or portion of a sequence) of a gene has been isolated, this sequence can be used to identify the entire gene, analyze the gene for homology to other sequences (i.e., identify it as a member of a gene family such as EGF receptor family) and then map the location of the gene on a chromosome. Accordingly, nucleic acid molecules described herein or fragments thereof, can be used to map the location of the gene on a chromosome. The mapping of the sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, genes can be mapped to chromosomes by preparing PCR primers from the sequence of a gene of the invention. These primers can then be used for PCR screening of somatic cell hybrids containing individual chromosomes. Only those hybrids containing the human gene corresponding to the gene sequences will yield an amplified fragment (For review of this technique se D'Eustachio, P. and Ruddle, F.H. (1983) "Somatic cell genetics and gene families." *Science* 220:919-924). Alternative methods of mapping a gene to its chromosome include in situ hybridization (see, for example, Fan, Y.S. et al. (1990) "Mapping small DNA sequences by fluorescence in situ hybridization directly on banded metaphase chromosomes." *PNAS* 87:6223-27), pre-screening with labeled flow sorted chromosomes (CITE), and pre-selection by hybridization to chromosome specific cDNA libraries. Furthermore, fluorescence in situ hybridization (FISH) of a DNA sequence to a metaphase chromosome spread can further be used to provide a precise chromosomal location in one step (see "Human Chromosomes: A Manual of Basic Techniques", Pergamon Press, New York, 1988). Lastly, with the completion (in the not-to-distant future) of the sequencing of the human genome, chromosome mapping will very quickly switch from elaborate, hands-on methods of mapping genes, to simple database searches

Once the sequence (or portion of a sequence) of a gene has been isolated, these agents can be used to assess the intactness or functionality of a particular gene. Comparison of affected and unaffected individuals can begin with looking for structural alterations in the chromosomes such as deletions, inversions, or translocations that are based on that DNA sequence. Once this is accomplished, the physical position of the sequence on the chromosome can be correlated with genetic data map. (such data are found, for example in McKusick, V. "Mendialian Inheritance in Man" available on-line through John Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in e.g. Egeland, J.A. et al. (1987) "Bipolar affective disorders linked to DNA markers on chromosome 11." *Nature*, 325:783-787). Alternatively, polynucleotide sequences can be used as probes in Southern Blot analysis to identify alterations in the organization of the gene of interest and surrounding regions. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms. If a specific mutation is observed in some or all individuals affected by a particular disease, but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease.

b. Tissue Typing

The nucleic acid sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the present invention are useful as additional DNA markers for RFLP mapping (described in US Patent 5,272,057). Furthermore the sequences of the present invention can be used to determine the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the nucleic acid sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the individual's DNA and subsequently sequence it. Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide

unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic variation. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The nucleic acid sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the non-coding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per 500 bases. Thus, each of the sequences described herein may be, to some degree, used as a standard against which DNA from an individual can be compared for identification purposes.

c. Forensic Biology

In addition the sequences described herein can be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example a perpetrator of a crime. To make such an identification, PCR-based technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, (e.g. hair, skin, or body fluids). The amplified sequence can then be compared to a standard thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents (e.g. PCR primers) targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual. The nucleic acid sequences described herein can further be used to provide polynucleotide reagents e.g. labeled or labelable probes, which can be used in, for example, an in situ hybridization technique, to identify a specific tissue. This technique can be exceedingly useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such probes can be used to identify tissue by species and/or organ type.

O. Predictive Medicine

Portions or fragments of the polynucleotide sequences of the invention can be used for predictive purposes to thereby treat an individual prophylactically.

unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic variation. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The nucleic acid sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the non-coding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per 500 bases. Thus, each of the sequences described herein may be, to some degree, used as a standard against which DNA from an individual can be compared for identification purposes.

c. Forensic Biology

In addition the sequences described herein can be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example a perpetrator of a crime. To make such an identification, PCR-based technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, (e.g. hair, skin, or body fluids). The amplified sequence can then be compared to a standard thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents (e.g. PCR primers) targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual. The nucleic acid sequences described herein can further be used to provide polynucleotide reagents e.g. labeled or labelable probes, which can be used in, for example, an in situ hybridization technique, to identify a specific tissue. This technique can be exceedingly useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such probes can be used to identify tissue by species and/or organ type.

O. Predictive Medicine

Portions or fragments of the polynucleotide sequences of the invention can be used for predictive purposes to thereby treat an individual prophylactically.

1. Diagnostic /Prognostic Assays

One method of detecting the presence or absence of a polypeptide or nucleic acid in a biological sample is to expose that sample to an agent that recognizes the entity in question. A preferred agent for detecting mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to the sequence one is attempting to detect (for instance, the sequence of the invention). The nucleic acid probe can be, for example, a full length cDNA, or a portion thereof such as an oligonucleotide of at least 15, 30, 50, 100, 250, or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to a mRNA or genomic DNA encoding the invention. The term "labeled" in this context refers to modifications in said sequences including, but not limited to, biotin labeling that can then be detected with a fluorescently labeled streptavidin, or ^{32}P labeling.

A preferred agent for detecting a polypeptide of the invention is an antibody or peptide capable of binding to the invention, preferably an antibody with a detectable label. Antibodies can be polyclonal or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g. a Fab or $F(ab)_2$) can be used. The term "labeled" in this context refers to direct labeling of the probe or antibody by coupling (i.e. physical linking) a detectable substance to the probe or antibody, such as a fluorescent labeled moiety or biotin.

The detection methods of the invention can be used to detect mRNA, protein, or genomic DNA in a biological sample in vitro as well as in vivo. For example, in vitro techniques for detection of mRNA include (but are not limited to) Northern Blot hybridization and in situ hybridizations. In vitro techniques for detection of a polypeptide of the invention include enzyme linked immunosorbent assays (ELISA's), Western blots, immunoprecipitations, and immunofluorescence.

The invention also encompasses kits for detecting the presence of a polypeptide or nucleic acid of the invention in a biological sample. Such kits can be used to determine if a subject is suffering from or is at increased risk of developing a disorder associate with aberrant expression of a polypeptide or polynucleotide of the invention. For instance, the kit can comprise a labeled compound or agent (as well as all the necessary supplementary agents needed for signal detection e.g. buffers, substrates, etc...) capable of detecting the

polypeptide, or mRNA in the sample (e.g. an antibody which binds the polypeptide or a oligonucleotide probe that binds to DNA or mRNA encoding the polypeptide).

The methods of the invention can also be used to detect genetic lesions or mutations in a gene of the invention, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant expression or activity of an agent of the invention. In preferred embodiments, the methods include detecting the presence or absence of a genetic lesion or mutation characterized by at least one alteration affecting the integrity of the agent of the invention. For example, such genetic lesions or mutations can be detected by ascertaining the existence of at least one of: 1) a deletion of one or more nucleotides from a gene; 2) an addition of one or more nucleotides to a gene; 3) a substitution of one or more nucleotides of the gene; 4) a chromosomal rearrangement of the gene; 5) an alteration in the level of a messenger RNA transcript of the gene; 6) an aberrant modification of the gene, such as of the methylation pattern of the genomic DNA; 7) the presence of a non-wild type splicing pattern of a messenger RNA; 8) a non-wild type level of the protein encoded by the gene; 9) an allelic loss of the gene; and 10) an inappropriate post translational modification of the protein encoded by the gene. Many techniques can be used to detect lesions such as those described above. For instance, mutations in a selected gene from a sample can be identified by alterations in restriction enzyme cleavage patterns. In this procedure, sample and control DNA is isolated, digested with one or more restriction endonucleases, and fragment length sizes (determined by gel electrophoresis) are compared. Observable differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Additional techniques that can be applied to detecting mutations include, but are not limited to, detection based on direct sequencing, PCR-based detection of deletions, inversions, or translocations, detection based on mismatch cleavage reactions (Myers, R.M. et al. (1985) "Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes." *Science* 230:1242), and detection based on altered electrophoretic mobility (e.g. SSCP, see, for example, Orita, M. et al. (1989) "Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms." *PNAS* 86:2766).

2. Pharmacogenetics

Pharmacogenetics deals with clinically significant hereditary variation in the response to drugs due to altered drug disposition and altered action in affected persons (see Linder, M.W. et al. (1997) "Pharmacogenetics: a laboratory tool for optimizing therapeutic efficiency." *Clin Chem.* 43(2):254-266). In general, two types of pharmacogenetic conditions can be differentiated. There are genetic conditions transmitted as a single factor altering the way drugs act on the body, referred to as "altered drug action". Alternatively, there are genetic conditions transmitted as single factors altering the way the body acts on drugs (referred to as "altered drug metabolism"). These two conditions can occur either as rare defects, or as polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency is a common inherited enyzmopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (e.g. anti-malarials, sulfonamides etc.).

The activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (e.g. N-acetyltransferase 2 (NAT2) and cytochrome P450 enzymes (CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM which all lead to the absence of functional CYP2D6. Poor metabolizers of this sort quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, a PM will show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so-called ultra rapid metabolizer who do not respond to standard doses. Recently, the molecular basis of ultra rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus the in the context of pharmacogenetics, an agent of the invention can be used to determine or select appropriate agents for therapeutic prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individuals drug responsiveness phenotype.

3. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents that effect the expression or activity of a polypeptide or polynucleotide of the invention can be applied in clinical trials. For example, the effectiveness of a drug directed toward a target identified by the invention and intended to treat a particular ailment, can be monitored in clinical trials of subjects exhibiting said ailment by monitoring the level of gene expression of the target, activity of the target, or levels of the target of the invention. Thus in a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent by comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of the polypeptide or polynucleotide of the invention in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level or activity of said target of the invention in the post-administration samples, (v) comparing the level of said target of the invention in the post administration sample with levels in the pre-administration samples, and (vi) altering the administration of the agent to the subject accordingly.

EXAMPLES

The following examples are intended to further illustrate certain preferred embodiments of the invention, and are not limiting in nature.

EXAMPLE 1: CELL LINES

S4535/HEK293 cells were propagated as monolayers in DMEM media (Gibco BRL) supplemented with 10% FBS, L-Glutamine (2mM final), non-essential amino acids (1X), Sodium Pyruvate (1mM), 300 ng/ml puromycin (ICN Biomedicals; Costa Mesa,

CA) and 300 μg/ml neomycin (Life Technologies; Gaithersburg, MD). The colorectal adenocarcinoma HT-29 cells and DLD1-1 cells were both obtained from ATCC and grown in McCoy's 5A media (Gibco BRL) modified with 10% FBS (Hyclone; Logan, UT). The packaging cell line, gp293 (a gift from Dr. Inder Verma, Salk Institute, CA) was maintained in DMEM, 10% fetal calf serum, and 200 μg/ml blasticidin (ICN Biomedicals; Costa Mesa, CA). HUVECs (Human umbilical vein epithelial cells, Clonetics Walkersville, MD) were maintained in EBM-2 basal media supplemented with a Bulletkit (Clonetics). HMECs (Human mammary epithelial cells, Clonetics) were maintained in DFCI-1 media (see Band, V. et al. (1989) "Distinctive traits of normal and tumor-derived human mammary epithelial cells expressed in a medium that supports long-term growth of both cell types." *PNAS* 86:1249-53). SW620 cells were maintained in DMEM (Life Technologies) enriched with 10% fetal calf serum. Cultures were grown at 33°C or 37 °C (5% CO_2) in standard tissue culture flasks. In some instances, Pen/Strep (1X, 100ug/ml ea.) was added to the cultures to minimize the risk of bacterial contamination.

EXAMPLE 2: CONSTRUCTION OF THE TBE-2 REPORTER

The vector (pBV-LUC) carrying the β-catenin /Tcf responsive promoter element consisting of 4 tandem repeats of the TBE-2 cassette was provided by the B. Vogelstein, (Johns Hopkins University). To construct a retroviral reporter vector, pBV-LUC was digested with MluI/BglII to remove the TBE-2 promoter and blunted using Klenow fragment. Subsequently, this fragment was ligated into the filled ClaI/BamHI sites of pVT806. As a result of these procedures, the product (pVT806-TBE2-EGFP) contains a TBE-2 promoter that is operably linked to coding sequence of EGFP and can be introduced into HEK293 cells using standard retroviral techniques.

EXAMPLE 3: ISOLATION AND CONSTRUCTION OF WT AND MUTANT β-CATENIN EXPRESSION VECTORS

β-catenin was derived from PCR amplification of cDNA prepared in-house. Specifically, total RNA prepared from HEK293 or HT29 cells was used to construct cDNA by methods common to the field (SuperScript, LifeTechnologies). Subsequently the N-terminal and C-terminal halves of β-catenin were amplified separately using

catenin-specific primers containing flanking restriction sites. (N-terminal primers =OVT 1801 5' CGCGGATCCGGCTACTCAAGCTGATTTGATGGAG and 1802 5'AGTC GTGGAATGGCACCCTGCTCAC C-terminal primers 1803 5'CTCCACAACC TTTTATTACATCAAG and 1804 5'TCCCCCGGGGCCAATCACAATGCAAGT TCAGACA; PCR conditions, i) 5', 94°C; ii) 40'', 94°C; iii) 40'', 57°C; iv) 2', 72°C; v) 10', 72°C, cycle through steps ii-iv 30 times, TAQ polymerase). The PCR products were subsequently purified (Qiagen PCR Kit), digested with the appropriate enzymes, and ligated into pCMV-TAG-3A. After validating the sequence of each half of the gene, N- and C-terminal clones were spliced together to form a full-length cDNA (Figure 9).

In order to construct an allele of β-catenin that was insensitive to regulation by APC, a synthetic allele carrying a substitution of tyrosine for serine at position 45 (S45Y) was constructed using QuikChange (Stratagene, see Kunkel, T.A. (1985) "Rapid and efficient site-specific mutagenesis without phenotypic selection." PNAS 82:488). Specifically, the full-length β-catenin clone was combined with Pfu Turbo and two complementary oligonucleotides encoding the required nucleotide change (OVT 1831,1832). The sample was then subjected to temperature cycling to allow incorporation of the mutated primers followed by digestion with DpnI (37°C for 1 hr.) to cut the parental DNA template. The resulting nicked, double-stranded molecules were then transformed into bacteria (DH5α cells) for repair of the newly mutated vector. The full-length cDNA of β-catenin S45Ywas subsequently cloned into the SacII/ClaI sites of the retroviral vector pVT312 and prepared for packaging in 293gp cells.

EXAMPLE 4: ISOLATION AND CONSTRUCTION OF WT AND MUTANT TCF-4 EXPRESSION VECTORS

Using techniques similar to those described above, Tcf-4 was PCR amplified from a HEK293 cDNA library. Specifically, the N- and C-terminal halves of Tcf-4 were constructed separately using TCF-4 specific primers flanked with either BamHI/PstI or PstI/EcoRI restriction sites (OVT 1805 5' CGCGGATCCGATGCCGCAGCTGAA CGGCGGTGGAG, 1806 5' TCTACGTCTGCAGGTAAGTGTGGAGGTG GGTTTC, and 1807 5'CTTACCTGCAGACGTAGACCCCAAAACAGGA, 1808 5' CAGCGGAATTCACGACGCTAAAGCTATTCTA). The N-terminal PCR product

yielded two products of approximately 605 bp in size. Sequence analysis subsequently identified the lower of the two bands as being the true N-terminus of Tcf. In contrast, PCR reactions designed to amplify the C-terminal region of Tcf yielded a minimum of three products that were later identified as splice forms Tcf-4B, Tcf-4DE, and Tcf-4DIE. The PCR products of both reactions were subsequently purified (Qiagen PCR Kit), digested with the appropriate enzymes, and ligated into pCMV-Tag-3A. Individual clones were then sequenced to confirm identity and then spliced together to form a full-length cDNA using standard molecular biological techniques (Figure 10). Upon confirmation of intactness of each construct, the cDNA's were ligated into the SacII/ClaI sites of retroviral expression vector, pVT312.

To create N-terminal deletions of Tcf that were capable of blocking the activity of β-catenin S45Y, PCR primers that initiated amplification at internal sites were used to isolate truncated Tcf cDNA's. Specifically, OVT 1826, which annealed 90 nucleotides from the N-terminus of Tcf was used in conjunction with the T7 primer to amplify a cDNA that lacked the N-terminal-most 30 amino acids. The resulting fragment was then cloned into pVT312 using techniques common to the art. In addition, the deletion mutants were also fused in-frame to the C-terminus of either the Gal4 BD or dGFP cDNA. To accomplish this, oligos specific to either Gal4 (OVT 1845, 1846) or dGFP (OVT 1848 and 1849) were used to PCR amplify the correct product from pFA-cJun and pVT352.1 respectively (Cycling conditions 5', 94°C; 40'', 94°C; 40'', 58°C; 1', 72°C; (cycle 16-20 times) 10', 72°C). Each of the fragments was then gel purified, digested with SacII and ligated (T4 ligase) into the equivalent site of pVT 312 carrying the Tcf-4DI deletion mutant. Restriction digest mapping was then used to identify which of the clones contained the scaffolding molecule in the correct orientation..

EXAMPLE 5: SELECTION OF A REPORTER CELL LINE

As a first step in identifying a reporter cell line, each of the necessary vectors including i) pVT806 TBE2-EGFP and ii) pVT312- β-catenin S45Y were packaged in 293gp cells using variations of one of two methods. In the first technique, the two constructs are co-transduced with VSV-G envelope expression plasmid into 293gp packaging cells (gift of I. Verma, Salk Institute) using LIPOFECTAMINE (Life Technologies). To accomplish this, 3 x 10^6 cells of the packaging cell line (293gp) are

seeded into a T175 flask. On the following day, two tubes are prepared, one containing 15 ug of pVT806 TBE2-EGFP DNA (or pVT312- β-catenin S45Y DNA) plus 10 ug of envelope plasmid (pCMV-VSV.G-bpa) in 1.5 ml DMEM (serum free), and the second containing 100 ug of LIPOFECTAMINE in 1.5 ml DMEM (serum free). These tubes are incubated separately at room temperature for 30 minutes, mixed, and incubated for an additional 30 minutes. This cocktail is referred to as the "transfection mix." The 293gp cells are then gently washed with serum free media and exposed to 20 ml of the transfection mix for 4 hours at 37 °C. The overlying mixture is then removed, the cells are washed once in DMEM (containing 10% serum) and then cultured in the same media. After 72 hours at 37°C the media (now referred to as "viral supernatant") is collected, filtered through a 0.45μm filter and frozen at –80°C. When needed, it is possible to make a second collection of virus by adding 15mls of DMEM (10% serum) back to the cells and incubating a further 24 hours.

As an alternative methodology, retroviral DNA can be packaged using a technique that is referred to herein as the "$CaCl_2$ Method". In one variation of this method, 5 x 10^6 cells of the packaging cell line (293gp) are seeded into a 15cm^2 flask on Day 1. On the following day, the media is replaced with 22.5 mls of modified DMEM. Subsequently, a single tube carrying 22.5μg of either pVT806 TBE2-EGFP (or pVT312-β-catenin S45Y) and 22.5μg of envelope expression plasmid (pCMV-VSV.G-bpa) is brought to 400μl with dH_2O, and 100μl of $CaCl_2$ (2.5M) plus 500μl of BBS (drop-wise addition, 2x solution = 50mM, BES (N,N-bis(2-hydroxyethyl)-2-aminoethane-sulfonic acid), 280mM NaCl, 1.5mM Na_2HPO_4, pH 6.95) is added. After allowing this retroviral mixture to sit at room temperature for 5-10 minutes, it is introduced to the 293gp cells in a drop-wise fashion, and the cells are then incubated at 37°C (3% CO_2) for 16-24 hours. The media is then replaced and the cells are allowed to incubate for an additional 48-72 hours at 37°C. At that time, the media containing the viral particles is then collected, filtered through a 0.45μ filter and frozen down at –80°C. Retroviral supernatant can subsequently be thawed and used directly to infect the appropriate host cells.

Viral supernatants of both pVT806 TBE2-EGFP and pVT312- β-catenin S45Y were then used to transduce HEK293 cells (ATCC # CRL-1573). Following standard procedures common to the art, a population of roughly 1 X 10^7 HEK293 cells were transduced

sequentially with each of the retroviral supernatants for 24 hours, using a 20% vol/vol of retroviral supernatant to complete media (KBM catalogue no.CC3101, Clonetics) plus 2% FBS. The cells were then allowed to recover for 24 hours in complete media followed by incubation in media containing the appropriate selectable marker (puromycin (0.3 μg/ml) and neomycin (600 ug/ml) to select cells containing stable inserts of the two constructs.

Fluorescent activated cell sorting (FACS) was subsequently used to identify individual clones capable of responding to the activation of the β-catenin-Tcf pathway (Coulter EPICS Elite Cell Sorter using EXPO "Build" and EXPO "Analysis" software). In the initial starting population (F0), a bi-modal histogram containing two overlapping populations, one dim and one weakly fluorescent, were observed. (Figure 11). Using standard sorting procedures, cells exhibiting a weak-bright fluorescent profile were collected (referred to herein as F1). When this population was expanded and reexamined by FACS, the proportion of bright and dim cells was observed to be more heavily skewed toward a bright peak. The brightest 5.6% of this population was then collected and expanded (referred to herein as F2). Subsequent FACS analysis of the expanded F2 population showed that the bimodal nature of the histogram was nearly absent and that the vast majority of the cells were moderately to highly fluorescent. The top 11.9% of this population (referred to herein as F3) was then sorted, and plated in 96 well plates at low density to isolate individual clones with highly fluorescent properties. Forty-five clones isolated from these procedures were then expanded into 6 cm^2 plates for further analysis. Upon FACS analysis, 6 of these clones were observed to show sufficient levels of fluorescence to warrant further investigation into the possibility of their use as reporter constructs.

To identify a reporter clone that could be modulated by the action of perturbagens, the retroviral construct, pVT312-dGFP-ΔTcf-30, carrying the dominant negative form of Tcf 4DIE fused to dGFP, was introduced into a sample of cells taken from each of the six clones. After three days, representatives carrying all three constructs were analyzed by FACS to determine which of the clones were responsive to the dominant negative inhibitor. Clones that failed to respond to the presence of TcfΔ30 (i.e. those that remained brightly fluorescent) were discarded. One clone, S4535, exhibited

the sort of properties that are desirable in a reporter construct. In the absence of TcfΔ30, >95% of the S4535 cells fell into the bright peak (Figure 12). In contrast, when the S4535 clone was transfected with the dominant negative form of Tcf, a bimodal histogram containing near equal numbers of bright and dim cells was observed, suggesting that S4535 is responsive to the presence of agents that disrupt the β-catenin-Tcf-APC pathway. Subsequently, the original parental clone carrying the β-catenin S45Y and TBE2-EGFP constructs, was expanded for future perturbagen screens.

EXAMPLE 6: PREPARATION AND TRANSFER OF A cDNA LIBRARY

Using techniques that are familiar to individuals in the art, randomly primed cDNA libraries were used as a source of sequences encoding putative β-catenin/Tcf pathway blocking agents. As one non-limiting example of how to construct such a library, polyA mRNA derived from placental tissue was PCR amplified using a random 9-mer linked to a unique SfiI sequence ("SfiA"), followed by an additional set of nucleotides that is used later for library amplification (OVT 906: 5' ACTCTGGACTAG GCAGGTTCAGTGGCCATTATGGCC$(N)_9$). The product of this reaction was size selected (>400 base pairs) and subjected to RNAse A/H treatment to remove the original RNA template. The remaining single stranded DNA was then subjected to a second round of PCR using a random hexamer nucleotide sequence linked to a second unique SfiI sequence ("SfiB") which was again followed by an additional set of nucleotides for future library amplification: (OVT 908: 5' AAGCAGTGGTGTCAACGCAGTGAGGCC GAGGCGGCC $(N)_6$). The final product of this reaction, a double stranded cDNA, was blunted/filled with Klenow Fragment (New England BioLabs), size selected, PCR amplified (OVT 909: 5' ACTCTGGACTAGGCAGGTTCAGT and OVT 910: 5' AAGCAGTGGTGTCAACGCAGTGA), digested with SfiI (New England BioLabs), and inserted into a retroviral vector (pVT 352.1, pBabe). As a result of these procedures, the sequences encoding the perturbagens were inserted at the 3' end of the non-fluorescent variant of EGFP (dEGFP). Expression of the dEGFP-perturbagen fusion gene (as well as the neomycin resistance gene present in the retroviral vector) was driven by the 5' LTR of pBabe. The library (~12 x 10^6 in size) was then packaged in 293gp cells (laboratory of I. Verma, Salk Institute) and retroviral supernatant was generated.

EXAMPLE 7: SCREENING FOR PERTURBAGENS THAT INHIBIT THE ACTIVATION OF THE TBE-2-EGFP REPORTER

To identify perturbagens that disrupted the β-catenin /Tcf pathway, the placental-derived perturbagen expression library was introduced into 20 x 10^6 S4535 cells using standard retroviral transduction techniques. Subsequently, the cells were cultured for 5 days and sorted to identify individuals within the population that exhibited decreased levels of GFP expression. In the first round of sorting, 14 x 10^6 cells of the population were collected. Genomic DNA was prepared from these cells and the perturbagen encoding insert was PCR amplified for sublibrary preparation. Specifically, the DNA encoding the perturbagens was PCR amplified from genomic DNA using two oligonucleotides that contained homology with sequences flanking the cDNA insertion site (oVT 181: 5' GGATCACTCTCGG CATGGACGAG and oVT 178: 5' ATTTTATCGATGTTAGCTTGGCCATT). One microgram of genomic DNA was added to each PCR reaction (along with the appropriate reagents to give 1x PCR cocktail, e.g. 2.5mM $MgCl_2$, 10mM oligos and 0.5-1mM dNTP and (Taq) polymerase (Boehinger Mannheim). Total PCR reaction volumes varied between 20 and 50ul/reactions and PCR cycling conditions followed the protocol of: 94°C, 2 minutes; 94°C 15 seconds; 68°C, 2 min and 30 sec (cycle 24 x); 68°C 3 minutes. The PCR product was then purified (Qiagen PCR purification kit), digested with SfiI (New England Biologicals) and directionally ligated (T4 ligase, Boehinger Mannheim) into the original vector (pVT352.1). This material was then transformed into bacteria by electroporation (DH10B, Electromax, Gibco) and inoculated into 500 ml of LB-Amp media for selection and expansion of bacterial cells that contained a member of the sublibrary. The bacterial cells were harvested at log phase of cell growth, and prepared (Qiagen) for isolation/purification of large quantities of the cDNA sublibrary. Subsequently, this material was re-packaged in 293gp cells in preparation for subsequent rounds of cycling and enrichment in S4535.

Following the second round of cycling, the population that fell within the dim gate was collected, replated, and expanded. These cells were then sorted (without sublibrary construction) to facilitate further enrichment of the cell population containing

perturbagen sequences. Rounds three and four included sublibrary preparation and proceeded in a similar fashion to round one.

The fraction of HEK293 cells expressing lower levels of GFP changed dramatically over the course of the 4 rounds of cycling/enrichment. In the early rounds of screening, the number of dim cells observed in the S4535 library-containing populations mirrored the percentage observed in control studies (5-8%). In contrast, by the end of 4 rounds of cycling/enrichment procedures, the number of dim cells had grown to 34.4 % (2.7 x over the observed background levels). Nearly one thousand library clones were picked from the fourth round of cycling, and screened individually using a high throughput screening procedure, to identify i) perturbagens that were capable of suppressing the expression of GFP in S4535 cells, and ii) agents that exhibited cytostatic or cytotoxic properties in secondary cell lines.

EXAMPLE 8: HIGH THROUGHPUT SCREENING

1. ISOLATION OF INDIVIDUAL PERTURBAGEN CLONES FOR HIGH-THROUGHPUT SCREENS.

Following multiple successive rounds of en-masse screening, the perturbagen encoding sequences were PCR amplified from the resultant sublibrary, recloned *en masse* into the appropriate retroviral vector (e.g. pVT 352.1 or pVT1515) transformed into bacteria (DH10B Electromax, Gibco) by electroporation, and plated on selective agar plates (LB, Amp^+) to identify individual clones that contained potential perturbagen inserts. To prepare purified plasmids of individual perturbagen clones, bacterial colonies were picked (either by hand or automated techniques, AutogenSys, Autogen) and grown overnight in a 96 well plate format (LB, +Amp media, 37°C). Samples from each well were then removed and frozen down as glycerol stocks which were later thawed, grown in liquid culture, and processed for plasmid DNA preparation.(Multiscreen Filtration Plates, Millipore).

2. HT PREPARATION OF VIRAL SUPERNATANTS AND TRANSDUCTION PROCEDURES

To obtain viral supernatants for HT screening procedures, 2 x 10^5 early passage 293gp cells (in 180 ul of media) were plated into each well of a 96 well microtiter dish

using either automated (Sorval "Cytomat" and a Beckman "Multimek" instrumentation) or non-automated techniques. The cells were incubated overnight to allow attachment to the solid support and then transfected with the individual retroviral miniprep DNA's. Several viable methods can be used to transfect cells in this format (e.g $CaCl_2$ .Lipofectamine, Transit™). In the $CaCl_2$ method, 133ng of library plasmid DNA was mixed with 534ng of envelope plasmid in a total volume of 5µl. $CaCl_2$ was then added (5µl) to a final concentration of 250 mM, followed by addition of an equal volume (10µl) of 2x BBS (50mM BES(N,N-bis(2-hydroxyethyl)-2-aminothane-sulfonic acid), 280 mM NaCl, 1.5 mM Na_2HPO_4, pH6.95). The solution was mixed every 5 minutes for 20 minutes before adding 20µl dropwise to the wells containing the gp293 cells. The cells were allowed to incubate 16 hours at 37°C, and the media was replaced with 100µl of fresh media. At 72 hours post transfection, the supernatant was removed to an empty 96 well plate, exposed to multiple freeze/thaw cycles (or filtration) to remove potential contaminant 293gp cells, and then frozen at –80°C for storage.

Transductions were performed by plating each cell line (e.g. S4535, HT29) in a microtiter plate in a total volume of 100 ul media. The following day, each retroviral supernatant was thawed, filtered through a 0.45 um Multiscreen-HV sterile filter plate (Millipore Corp., Bedford, MA) and added to the cells along with polybrene (4ug/ml). In most instances, the viral supernatant represented 50% of the volume of the final mixture.

3. HT TransFACS Screen for Modulators of TBE2-GFP Expression in HEK293 Cells.

To test the ability of individual perturbagens to down-regulate the TBE2-EGFP reporter construct, three hundred S4535 cells were plated in each well of a 96-well format and allowed to attach overnight. Subsequently, a single viral supernatant (85ul) and polybrene (final concentration of 4ug/ml) was added to each well and allowed to incubate for 16 hours. Following viral infection, the media was replaced with 200ul of fresh culture media. All of the above mentioned procedures were performed by hand or made use of the Sorval "Cytomat" and a Beckman "Multimek" instrumentation. Perturbagen expressing S4535 cells were cultured for 6 days at 37°C (5% CO_2) before being analyzed by FACS. To prepare the cells for analysis, each well was washed 1x with PBS and then treated with trypsin (50ul of a 0.05% solution. + 53mM EDTA, 10

minutes, 37°C, Life Technologies; Gaithersburg, MD) to release the cells from the surface of the well. Subsequently, 150 ul of DMEM + 10% FCS was added to each well to neutralize the trypsin and samples were then analyzed on the FL1 channel of a Coulter Epics XL-MCL (Beckman Coulter; Fullerton, CA) using EXPO software and an automated 32-position sample carousel.

Using the procedures described above, nine hundred and fifty-seven clones (obtained after four rounds of TransFACS screening, en mass) were tested for their ability to alter GFP expression levels in the S4535 clonal line. Of the roughly 1,000 clones analyzed, roughly 8% (76 clones) were judged to be positive (i.e. the perturbagen inhibited expression of GFP) based on positive and negative controls previously described (e.g. pVT1515 and pVT312-dGFP-ΔTcf-30). When the DNA sequences of these clones were matched with the FACS data, all the positive clones were found to be fragments of various forms of cadherin including cad5, cad11, and cadherin 6 (a member of the type II cadherins, see Shimoyama, Y. et al. (1995) "Isolation and sequence analysis of human cadherin-6 complementary DNA for the full coding sequence and its expression in human carcinoma cells." *Cancer Res.* 15;55(10):2206-11). The sequences of the cadherin clones selected in the β-catenin assay were similar (see Figures 13-15). All comprised C-terminal fragments representing the cytoplasmic domains of the native molecules fused to GFP. The smallest fragment, GFP-cad11, was 93 amino acids (not including GFP). Each cadherin sequence included the region known to interact with β-catenin (Huber and Weis, 2001) and exhibited a penetrance that was similar to TCF4Δ30 (see Figure 16).

4. Testing Cad5 C-terminal Fragment for Differential Cytostatic and Cytotoxic Activity using a HT assay

The cytotoxic and cytostatic activity of Cad perturbagens was assayed using a variation of the high-throughput techniques described in US No. 60/305,712, VEN012/00, "Automated Assay Methodology," the contents of which are incorporated in this document, in full. The procedures involved in this assay include i) transducing perturbagen encoding retroviral constructs into a suitable cell line(s) in a 96-well format, ii) culturing cells for a sufficient length of time necessary to observe the phenotype of

interest, and iii) assaying said cells for phenotype of interest (i.e. cell death and total cell number).

In one non-limiting technique, retroviral transductions were performed by plating roughly 10^3 cells of each line in a microtiter well of a 96 well plate (total volume = 100 µl media) and allowing the cells to attach over the course of several hours (or overnight). Subsequently, the retroviral supernatants were added to the cells along with polybrene (4µl/ml). In most instances, the viral supernatant represented 50% of the volume of the final mixture. In another transduction format, 84µl of filtered viral supernatant was combined with 100µl of cells seeded the previous day in black plastic walled, clear bottom 96-well plate(s). The seeding density of the plates varied, depending on the cell type, and was determined empirically as the density that would produce an approximately 80% confluent culture by six days post-transduction. The number of seeded cells were as follows: HT29 = 500 cells/well; HEK293 = 300 cells/well; primary HMEC = 2,000 cells/well; primary HUVEC = 200-1000 cells/well, depending on the passage. Polybrene was also added to the transduction to a final concentration of 4µg/ml. Approximately 16 hours post-transduction, the media was changed, and incubation was allowed to continue at 37°C.

To determine the cytotoxic/cytostatic effects of a given perturbagen, cells transduced with the agent of choice were analyzed to determine the number of dead and live cells remaining in the well after a given period of incubation. There are several methods that can be used to quantitate the number of dead and/or dying cells. As one non-limiting example, the ensuing procedures are followed: five days post-transduction, Sytox Orange (Molecular Probes, Eugene, OR) is added to each well of the assay plate to a final concentration of 1µM. The plates are then allowed to incubate for 30 minutes at 37°C, and then analyzed on a CCD imaging system (Sytox Orange, Ex:535+/-15nm, Em 585 +/-20nm) to determine the number of cells having a compromised cell membrane (i.e. dead and/or dying cells). In this instance, the imaging system is composed of a PixelVision Spectra Video™ Series imaging camera (1100 x 330 back-illuminated array, Pixel Vision, Tigurd, OR), Pixel Vision PixelView™ 3.03 software, two 50mm/f2 Olympus macro focusing lens mounted front to front, four 20750 Fostec xenon light sources (Schott-Fostec, Auburn, NY), four 8589 Fostec light lines, a 4457 Daedal stage,

and supporting mechanical fixtures. Mechanical fixtures were constructed to position the PixelVision camera below a microtiter dish. Additionally, the fixtures mounted the four Fostec light lines and allowed the excitation light to be focused on the viewed area of the microtiter dish. A 510 nm filter was placed between the two lenses. The front-to-front lens configuration provides 1:1 magnification and close placement of the target object to the imaging system.

After the dead-cell count is executed, a total cell number is determined by adding a detergent, e.g. saponin or tween-20, to each well to permeablize the remaining cells (0.1%, 30 minutes). Subsequently, an additional readout is performed to determine the total well fluorescence, (an indirect measure of the total number of cells). The number of dead and live cells in each well is then compared with the appropriate controls to determine if any cytotoxic/cytostatic properties are associated with the agent under study.

To investigate the effects of the Cad5 C-terminal fragment on cell viability and growth, a pVT1515-dGFP-Cad5 plasmid was introduced into four separate cell types (HT29, HEK293, HuVEC, and HMEC) using the techniques described above. The HT29 colon carcinoma cell line contains an inactivating mutation in APC (and hence a constitutively activated β-catenin pathway). In contrast, the transformed kidney cell line, HEK293, and two primary cultures (HMECs and HUVECs) contain a wild type (regulated) β-catenin pathway. In concurrent, parallel experiments, the parent plasmid, pVT1515, and a positive control plasmid (pTcfDN) were also transduced into the test cells. Measurements of dead cell numbers and total cell fluorescence were then collected on all samples using the Sytox staining procedure. In a single experiment, a microtiter plate with 8 replicates per clone was used, and multiple plates were analyzed. In all, between 24 and 97 replicates of each construct were examined.

Looking at cell death as the first key parameter, analysis show that like TcfDN, the C-terminal fragment of Cad5 exhibits a strong cytotoxic effect in HT29 cells (Figure 17). These results are in contrast with the observations made in cell lines carrying an intact beta-catenin pathway. In HEK 293 cells, HMECs, and HUVECs, the cytotoxic effects of the Cad5 perturbagen are considerably smaller and approach the numbers induced by introduction of the parental plasmid, pVT1515. Analysis of total cell numbers in each of the cell lines further supports the notion that the C-terminal fragment of Cad5 has a more

deleterious effect in cells containing a disrupted β-catenin/APC pathway (Figure 18). The over-expression of the Cad5 perturbagen caused a significant reduction in the number of HT29 cells, yet did not substantially alter the growth of the other cells tested. This is in contrast to the dominant negative Tcf construct that did not make an observable alteration in the numbers in HT29 cells, but decreased cell numbers in all non-colon lines. These data demonstrate that the Cad5 C-terminal fragment embodies both cytotoxic and cytostatic properties. Furthermore, the effects of Cad 5 are accentuated in colon cancer cells containing defects in the β-catenin/APC pathway, while exhibiting less noticeable consequences in other tissues having an intact pathway.

5. Targets of the Cad5 Perturbagen

Microarray expression studies were used to identify additional targets of the Cad5 perturbagen. To determine which genes were induced (or repressed) in cells containing the Cad5 C-terminal fragment, the Cad5 agent was introduced into S4535 cells, and polyA RNA was isolated and readied as probes for microarrays. To accomplish this, S4535 cells were seeded into 96 well tissue culture plates (Becton-Dickinson; Bedford, MA) at 300 cells/well (in a total of 100μl) one day prior to transduction. On the day of transduction, Cad5, TcfDN, or control (pVT1515) viral supernatants were thawed, and 85 μl of supernatant was added to the wells containing S4535 cells. Polybrene was added to 4μg/ml to enhance infection. Sixteen hours post transduction, the media was changed, and the cells were allowed to continue incubation until 6 days post transduction.

At six days post transduction, the media was removed, the cells were washed with PBS, and 50μl of a solution containing 0.05% trypsin plus 0.53mM EDTA (Life Technologies; Gaithersburg, MD) was added to each well. After ten minutes, the trypsin was neutralized by the addition of 150μl of DMEM with 10% fetal calf serum and a portion of the suspension was analyzed by fluorescent activated cell sorting (FACS, Coulter Epics XL-MCL; Beckman Coulter; Fullerton, CA) using an automated 32-position sample carousel. Results of these procedures demonstrated that the large majority (>95%, see Figure 19) of S4535 cells carrying either the Cad5 (or TcfDN) fragments had switched off reporter expression, thus confirming expression of the inhibitor. Total RNA and polyA RNA was then prepared from roughly 1 x 10^8 cells taken from each group (Rneasy, Oligotex, Qiagen; Valencia, CA). The intactness of each sample was then inspected by agarose gel

electrophoresis and quantitated by spectrophotometer prior to performing differential display analysis (Incyte Genomics, St.Louis, MO).

To analyze the data from the roughly 9,000 human genes and ESTs that were present on each microarray slide, the statistical analysis proceeded using the method of Kamb, A., "A simple method for statistical analysis of intensity differences in microarray-derived gene expression data" (submitted for publication). Briefly, the internal controls accompanying the Incyte data files were removed to allow data manipulation. Intensity differences (*S1-S2*), and average signals ((*S1+S2*)/2)) for the pVT1515 control RNA (in separate hybridizations) were then calculated for each analogous gene pair on the slide. These columns were then sorted (in descending order) based on the average signal and averaging window sizes were tested. After settling on 100 data points as the window, an averaged (*S1+S2*)/2 incremented by one point each time was calculated along with σ and σ^2 values for the corresponding sets of 100 points in the intensity column. Polynomials and lines were fit to plots of avg. (*S1+S2*)/2 vs. $\sigma_{\Delta S}^2/2$. These fits were used to compute σ_S^2 values at each signal intensity. Sequences with very low average signals ((*S1+S2*)/2 < 400) corresponding to 20% of the dataset were removed. To calculate a *z* statistic, each ΔS value for the inhibitor RNA – control (pVT1515) was divided by its corresponding $\sigma_{\Delta S}$ (the distribution mean, μ, was approximately zero, and was not subtracted from ΔS to find *z*).

Analysis of the microarray experiments comparing Cad5CD and TcfDN revealed highly correlated datasets, suggesting that both agents affected S4535 cells in a similar fashion. Interestingly, though there were many statistically significant differences in expression (when compared with control studies) the largest single change was only 2.3-fold (the Dickkopf-1 gene) suggesting the phenotypic similarities between Cad5CD (and TcfDN) were related to a large number of relatively small alterations in gene expression. It should be noted, however, that dynamic range compression of microarray data, cell-to-cell variation in inhibitor expression level, and population averaging of the cellular response at the RNA level could cause underestimation of the ratios in cells and/or their biological significance.

A group of genes whose expression levels differed significantly from controls ($|z| > 3$; $p < 0.01$) were identified in Cad5-perturbagen containing cells. Interestingly, many

of these Cad5CD targets exhibited similar patterns of gene expression in both TcfDN-containing and Cad5CD-containing cells, suggesting a common effect of Cad5CD and TcfDN on gene expression in HEK293 cells (see Figure 20). Still other genes were found to be modulated only in cells carrying the Cad5CD perturbagen (see Figure 21). Amongst the genes identified were several growth-related genes including an NDP kinase, and tyrosine phosphatase epsilon, both of which have been implicated in tumor formation (Elson, A. (1999) "Protein tyrosine phosphatase epsilon increases the risk of mammary hyperplasia and mammary tumors in transgenic mice." *Oncogene, 18:* 7535-42; and Nakayama, T., et al. (1992) "Expression in human hepatocellular carcinoma of nucleoside diphosphate kinase, a homologue of the nm23 gene product." *J. Natl. Cancer* Inst., *84:* 1349-54). In addition two genes related to interferon response were found to be modulated. A transforming sequence from adenovirus (E1b) was also present among the repressed genes in both datasets, a potentially significant observation given the role of adenovirus in creating the HEK293 cell line (Nakayama, T. et al. (1992). "Expression in human hepatocellular carcinoma of nucleoside diphosphate kinase, a homologue of the nm23 gene product. " *J. Natl. Cancer Inst., 84:* 1349-54).

Genes with known roles in the β-catenin pathway were also observed amongst the target group. Importantly, cyclin D1—known to be repressed by TcfDN—was in the category of down-regulated genes in Cad5 containing lines. In addition, axin-2 (also known as conductin), which acts upstream in the β-catenin pathway to inhibit β-catenin function (Nakamura, T., et al. (1998) "Axin, an inhibitor of the Wnt signalling pathway, interacts with beta-catenin, GSK-3beta and APC and reduces the beta-catenin level." *Genes Cells, 3:* 395-403; Kishida, S. et al. (1998) "Axin, a negative regulator of the wnt signaling pathway, directly interacts with adenomatous polyposis coli and regulates the stabilization of beta-catenin" *J. Biol. Chem., 273:* 10823-6.), was down-regulated, suggesting the existence of a compensatory feedback mechanism in the β-catenin pathway as a consequence of Cad5CD (and to a lesser extent TcfDN) inhibition. Cadherin-2, another negative regulator, was also repressed by Cad5CD.

As is apparent to one of skill in the art, various modifications of the above embodiments can be made without departing from the spirit and scope of this invention. These modifications and variations are within the scope of this invention.